JOURNAL OF APPLIED PHYSICS

JOURNAL
OF
APPLIED PHYSICS

Vol. 23, No. 9 SEPTEMBER, 1952

In This Issue

		Page
Markings Found in the Oxide Replicas for Electron Microscopy	Shigeto Yamaguchi	935
Study of the Oxide-Coated Cathodes by X-Ray Diffraction Method (I)	Eiso Yamaka	937
Turbulent Diffusion and Erosion	Richard W. Davies	941
Radiation from a Dielectric Wave Guide	Robert M. Whitmer	949
Dimensional Changes Normal to the Direction of Diffusion	R. W. Balluffi and B. H. Alexander	953
On the Metallic Shadow-Casting Using a Nozzle System	Tadatosi Hibi	957
Reflection of Microwaves from Metal-Plate Media	M. D. Pearson, and S. R. Peoples	964
Method for Determining Magnetic Moments and for **Measuring Susceptibilities and Permeabilities**	S. J. Barnett	975
Temperature Dependence of the Viscosity of Liquids	F. Gutmann and L. M. Simmons	977
On the Calorimetric Heat Exchange Correction	H. D. Vasileff	979
Statistical Approach to the Space-Charge Distribution **in a Cut-Off Magnetron**	Gunnar Hok	983
Determination of Complex Dielectric Constants of Absorptive Liquids by Microwave Interferometry	Franklin H. Branin, Jr.	990
Propagation of Elastic Waves in a Cylindrical Bore Containing a Fluid	M. A. Biot	997
Conductivity Change in Polyethylene During γ-Irradiation	S. Mayburg and W. L. Lawrence	1006
Metal Transfer and Wear	I.-Ming Feng	1011
Treatment of the Viscosity of Concentrated Suspensions	Robert Simha	1020
Effect of Accelerating Voltage and Specimen Morphology on Electron Diffraction Patterns	S. G. Ellis	1024
Methods of Measuring the Properties of Ionized Gases at High Frequencies. III. Measurement of Discharge Admittance and Electron Density	David J. Rose and Sanborn C. Brown	1028
Self-Diffusion in Pure Polycrystalline Silver	L. Slifkin, D. Lazarus, and T. Tomizuka	1032
Reliability of Probe Measurements in Hot Cathode Gas Diodes	Gottfried Wehner and Gustav Medicus	1035
On the Theory of Prediction of Nonstationary Stochastic Processes	R. C. Davis	1047
Letters to the Editor:		
Angular Shift of the Transmitted X-Ray Beam near the Bragg Angle	L. Carlton Brown, George L. Rogosa, and Guenter Schwarz	1054
Shadow of an Electron Beam	H. M. Smith and S. Hansen	1054
Self-Heating Thermionic Vacuum Tubes	E. G. Hopkins	1055
Kirkendall Effect in Alloy Systems	Herbert N. Hersh	1055
Attempt to Explain a Discrepancy in Observations of the Heat Transfer between a Vertical Surface and a Fluidized Bed of Particles	Max Jakob	1056
Some Variational Principles for Problems in Transmission Lines	Mikio Namiki and Hiroshi Takahashi	1056
Metallic Corrosion Influenced by Ultrasonic Waves	Shigeto Yamaguchi	1057
Significance of Flow-Patterns for Initial Convection in Porous Media	H. L. Morrison and F. T. Rogers, Jr.	1058
Separation of Stacking Fault Broadening in Cold-Worked Metals	B. E. Warren and B. L. Averbach	1059
Books Reviewed		1060
Other Books Received		1062
Announcements		1064
Index to Advertisers		xxx

BOARD OF EDITORS
Elmer Hutchisson,
Editor
Earle C. Gregg,
Assistant Editor

Associate Editors
Howard H. Aiken
Ray E. Bolz
Lyle D. Borst
Robert D. Heidenreich
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Roman Smoluchowski

Publication of the
American Institute of Physics

OFFICERS
George R. Harrison,
Chairman
George B. Pegram,
Treasurer

GOVERNING BOARD
Karl K. Darrow
G. J. Dienes
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Hugh S. Knowles
C. C. Lauritsen
F. W. Loomis
William F. Meggers
George B. Pegram
I. I. Rabi
Duane Roller
Frederick Seitz
Francis G. Slack
John C. Steinberg
Richard M. Sutton
Mark W. Zemansky

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton,
Director
Wallace Waterfall,
Executive Secretary
Theodore Vorburger,
Advertising Manager
Ruth E. Bryans,
Publication Manager
Betty H. Goodfriend,
Assistant Secretary
Eileen B. Neuberger,
Circulation Manager
Kathryn Setze
Assistant Treasurer

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 23, Number 9 September, 1952

Markings Found in the Oxide Replicas for Electron Microscopy

SHIGETO YAMAGUCHI
Scientific Research Institute, Ltd., Komagome Bunkyo-ku, Tokyo, Japan
(Received April 17, 1952)

There are two types of crystal edges in the electron micrographs of oxide replicas of metals. The first type is characterized by the edges being more transparent than the facets in reprinted crystals. In the second type of reprinted crystal there were found the opaque edges of crystal. These characteristics of crystal edges were explained by the orientation of crystallites composing the replica films. It is proposed in the present study that the width of transparent edge of reprinted crystal might serve as a rough measure of microscopic magnification.

THE oxide films formed on metals and alloys can be used as replicas for electron microscopy. The bright-illuminated micrographs of these oxide replicas are classified into two types. The first type has a marking in which the edges of the reprinted crystals are more transparent or thinner toward electrons than the other domains of oxide film corresponding to the facets. The second type is characterized by a marking wherein the crystal edges in replica are more opaque or thicker than the facets.

EXPERIMENTAL RESULTS

The oxide replicas with a marking of the first type were given by the oxide films composed of oriented crystallites. The second type was given by the non-oriented replicas. This was derived from the following experimental facts.

1. It is discernible in the micrographs obtained by the author[1] that the edges of octahedrons of reprinted crystals, which were from an etched face of iron single crystal, are transparent and those from Permalloy (Ni: 81.5 percent) single crystal are opaque to electrons (see Figs. 1(A) and (B)). A diffraction pattern obtained with a beam passing through Fe_3O_4 film of Fig. 1(A) was regularly oriented, whereas that of Fig. 1(B) consisted of the Debye-Scherrer rings verifying the existence of the NiO crystallites of random arrangement.[1,2]

2. We have the micrographs of transparent crystal edges, which were obtained from the oxide replicas of etched aluminum (see the micrographs obtained by Mahl,[3] Grivet,[4] and Castaing[5]). The oxide replica of

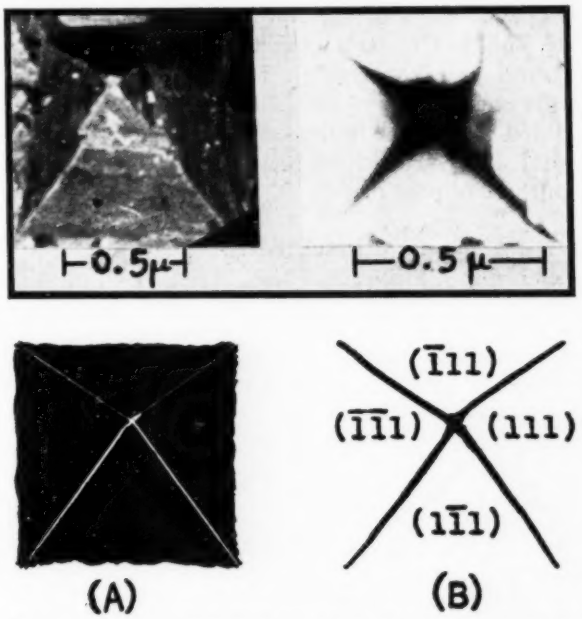


FIG. 1. (A) An octahedral crystal with transparent edges in oxide replica. (B) An octahedral crystal with opaque edges in oxide replica. (A) was obtained with the Fe_3O_4 film formed on iron. (B) was obtained with the NiO film formed on Permalloy.

[1] S. Yamaguchi, J. Appl. Phys. **22**, 1295 (1951).
[2] S. Yamaguchi, J. Appl. Phys. **23**, 445 (1952).
[3] H. Mahl, "Metallurgical Applications of the Electron Microscope," The Institute of Metals, London, Monograph and Report Series No. 8, p. 37 (1950).
[4] P. Grivet, reference 3, p. 19.
[5] M. R. Castaing, reference 3, p. 156.

FIG. 2. (A) An oxide film with transparent edge composed of regularly oriented crystallites. (B) An oxide film with opaque edge composed of crystallites arranged at random.

aluminum etched with aqua regia shows the transparent edges of reprinted crystals, whereas the oxide film obtained from aluminum etched *twice* with aqua regia shows the opaque edges of crystals, which have been made somewhat round by the repeated etching (see a micrograph obtained by Grivet[4]). The former film evidently is composed of crystallites oriented regularly according to the orientation of the underlying substrate. The latter film must consist of the crystallites arranged at random, because the underlying substrate, which has been repeatedly etched, does not possess crystal facets able to determine the orientation of oxide film formed on it.

3. The oxide replica obtained by the aluminum-pressing method using an amorphous surface layer shows the opaque edges of reprinted crystals (see a micrograph obtained by Nutting and Cosslett[6]). It is obvious that the oxide replica prepared by this method is never composed of oriented crystallites.

4. The oxide replicas obtained from alloys often give the opaque edges of crystals (see the micrographs[7] of 18–8 and 13 Cr steels obtained by Mahla, Nielsen, Katsurai, and Yamaguchi). The mixed oxides formed on the etched faces of alloys are hardly ever regularly oriented. This is due to the same causes as for the single crystal of alloys, which are prepared with more difficulty than those of pure metals.

DISCUSSION

Figures 2(A) and (B) are models of the oriented and non-oriented films, respectively. Figure 2(A) gives rise to a transparent crystal edge in micrograph. Figure 2(B) gives rise to an opaque edge. The model of Fig. 2(A) is in accordance with the experimental fact that the oriented Fe_3O_4 film formed on iron single crystal is very fragile as compared with that of Fig. 2(B) formed on Permalloy, as has been already mentioned in a previous report.[2] Figure 2(A) means that the oriented oxide crystallites grow up in a direction perpendicular to the facets of the underlying substrate under the conditions of lattice fit.

The appearance of the transparent edges of the oriented film seems to be explained by the refraction of electrons by the medium with parallel faces. However, the effect of refraction is so slight here, that it can not give rise to the transparent edge whose width amounts to about 500A. This is due to the condition that the inner potential (10–20 volts) of the oriented oxide crystals with a thickness of 500–1000A can refract very little the electrons accelerated by a potential of about 50 kv, so far as the incident angle of beam to the film face is about 45 degrees.

TABLE I. This table shows that the widths of the transparent edges of crystals measured in the micrographs of oxide replicas are roughly constant (300–800A).

Metals	Width of transparent edge	Observed by
Al	500A	Grivet[a]
Al	800	Mahl[b]
Al(Mg:0.78, Si:0.5)	400	Castaing[c]
Al	300	Brown[d]
Fe	300	Yamaguchi[e]

[a] See reference 4.
[b] See reference 3.
[c] See reference 5.
[d] A. F. Brown, "Metallurgical applications of electron microscope" (The Institute of Metals, London, 1950), p. 103.
[e] See reference 1.

It is remarked in the micrographs obtained by several workers (see Table I) that the real widths of the transparent edges measured are roughly constant (300–800A) and comparable with the thickness of the oxide replica for electron microscopy.

CONCLUSION

1. The existence of the transparent edges of the reprinted crystals, which are observed in the electron micrographs of oxide replicas, can reveal that the oxide crystallites composing the film are oriented regularly according to the orientation of the underlying metal faces.

2. The widths of these transparent edges found in the oxide replicas might serve as a rough measure of microscopic magnification, because they are always approximately constant (300–800A).

[6] J. Nutting and V. E. Cosslett, reference 3, p. 57.
[7] E. M. Mahla and N. A. Nielsen, J. Appl. Phys. **19**, 378 (1948); T. Katsurai, Trans. Chalmers Univ. Tech., Gothenburg, Sweden, Nr. 96 (1950); S. Yamaguchi, Bull. Chem. Soc. Japan **24**, 237 (1951).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

A Study of the Oxide-Coated Cathodes by X-Ray Diffraction Method (I)

EISO YAMAKA
Electrical Communication Laboratory, Tokyo, Japan
(Received July 2, 1951)

The relation between the crystal size of barium-strontium oxide and its thermionic emission was studied with the same vacuum tube. The crystal size was measured by an x-ray diffraction method. The measurements were made in vacuum by using thin mica windows as the entrance and exit ports for the x-ray beam. The emission was measured with pulse in order to avoid gas from the anode. Below 1050°C the crystal grows slowly, but above it crystal growth is rapid. It seemed that impurity activation is dominant at low temperatures, but heat activation is dominant at high temperatures.

I. INTRODUCTION

IT is said that the crystal growth of barium oxide is more rapid than barium-strontium oxide produced by heating the cathode, therefore, the former is inferior to the latter as an emitter. On the contrary, it is said that barium-strontium-calcium oxide grows more slowly than barium-strontium oxide. However, the relations between the crystal growth of the oxide and its components, or the crystal size and its emission, have not yet been investigated owing to the difficulty of the measurement of crystal size in vacuum. Kawamura[1] and others reported the decrease of activation centers in the oxide-coated cathode as a result of crystal growth when water vapor or chlorine gas were introduced into the vacuum tube, but they did not measure the crystal size directly. Recently, Eisenstein[2] investigated in detail, by x-ray diffraction methods, the decomposition of the carbonate and the evaporation of the oxide from the surface by heat treatment of the cathode, but the emission was not measured simultaneously, the relation between the crystal size and the emission being left uninvestigated. In addition, he prepared the sample by cracking the test tube in dry nitrogen and dipping the cathode into protecting wax, rendering the repetition of measurements impossible.

To emphasize the merit of the repetition of measurements it is desirable that the measurement of the crystal size be performed without cracking. Recently, we have succeeded in making a vacuum tube with two mica windows with which the crystal size could be measured repeatedly by passing x-rays through the mica windows. Simultaneous measurements of the electron emission from the cathode could also be made with the movable anode.

II. EXPERIMENTAL PROCEDURE

Figure 1 shows a diode tube with two mica windows. A thin mica sheet, 3/100 mm in thickness and 20 mm in diameter, is attached to the edge of the vacuum tube by solder glass. The anode is constructed to be movable in order to measure both the crystal size and the emission within the same test tube. An ionization gauge designed by Fogel[3] is connected to the test tube for the measurement of vacuum. The temperature of the oxide is measured by a platinum and platinum-rhodium thermocouple.

First, the test tube is connected to a vacuum system and evacuated to 1×10^{-6} mm Hg. After baking-out for 2 hours at 300°C, out-gassing of the anode is carefully carried out by high frequency induction heating. Then the cathode is heated at about 820°–870°C, a temperature slightly above that of the decomposition of the carbonate. After flashing the tube-getter, the tube is sealed off from the vacuum system without activation of the cathode in order to avoid undesirable crystal growth by the heat treatment.

The crystal size of the oxide is measured by the x-ray diffraction method. As shown in Fig. 2, an x-ray beam of Cu-K radiation, which is narrowed by two slits with a 1-mm gap, is lead into the tube through one of the two mica windows and allowed to fall upon the cathode. The beam is diffracted by the cathode in the direction of 2θ, where θ is the Bragg angle. If the (111) line of the oxide is selected for the measurement, the beam diffracted by two nickel supporting rods for the anode does not enter the Geiger-Müller counter, even if the incident beam irradiates the rods, because fortunately 2θ for the oxide is about 30° in contrast to about 44° for the Ni rods. The intensity distribution of the diffracted beam is measured in the neighborhood of 2θ by the thin mica window Geiger-Müller counter

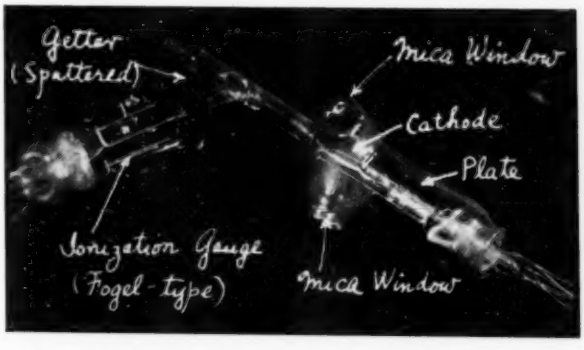


FIG. 1. Test tube with two mica windows as the entrance and exit ports for x-ray beam.

[1] H. Kawamura, J. Phys. Soc. Japan **3**, 301 (1948).
[2] A. Eisenstein, J. Appl. Phys. **17**, 654 (1946).
[3] Fogel, Proc. Inst. Radio Engrs. **5**, 302 (1946).

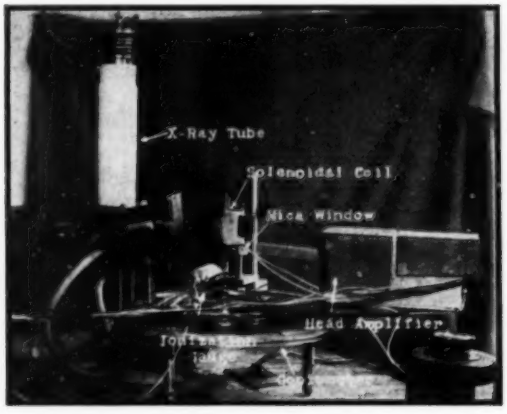


FIG. 2. Apparatus of crystal size measurement.

which is sensitive to x-ray, as shown in Fig. 3. The pulses from the counter are introduced into a scale-of-16. The crystal size of the oxide is calculated from the line broadening by the following relation;

$$D=K\lambda/B'\cos\theta$$
$$B'^2=B^2-b^2,$$

where D is the diameter of the crystal, K is a constant, B' is the true broadening of the diffracted beam, λ is the wavelength of x-rays used, θ is the Bragg angle, B is the observed broadening of the diffracted beam, and b is the instrumental broadening. b is measured using quartz powder much larger than 1000A. As we assume $K=1$, there will be about ± 10 percent error in the crystal size calculated from the above formula. Although the unmonochromized x-ray beam of Cu is used, the diffracted K_β beam is separated from the K_α beam about 3° in 2θ, therefore there is no possibility that the existence of the K_β beam makes a serious error in the determination of the line broadening. The error caused by the coexistence of $K_{\alpha 1}$, and $K_{\alpha 2}$ is negligibly small,[4] and we do not calibrate this error.

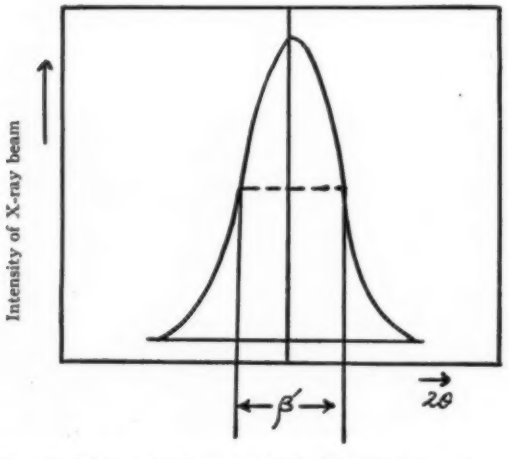


FIG. 3. Schematic shape of (111) diffracted x-ray beam and half-line breadth.

[4] L. Alexander and P. Klug, J. Appl. Phys. **21**, 137 (1950).

Thermionic emission from the oxide is measured with pulses in order to avoid gas from the anode which would be developed by static measurements. The measurement is carried out at a temperature lower than that used in processing the cathode, and the Richardson line is plotted. The peak value of the emission drawn on the oscilloscope is thought to be the temperature-saturation value of the emission.

The test tube shown in Fig. 1 is mounted inverted at the center of the goniometer throughout this experiment. The anode is moved up and down by switching on and off the current into solenoid coil, which covers the body of the tube. The anode is raised for the measurement of the crystal size by the x-ray beam, and is lowered for the measurement of the emission.

Following the heat treatment of the cathode for a definite time at a temperature, say T_1, higher than that of the decomposition of the carbonate, the measurements are carried out in the following order; measurements (1) of vacuum, (2) of the emission, and (3) of the crystal size. Next, at the same temperature, T_1, the cathode heating is repeated for a definite time and the measurements are carried out again. When the measurements are continued at T_1, the crystal size and its emission gradually saturate as shown in Fig. 4. After the saturation is reached at T_1, the measurements are carried out at a higher temperature T_2. Curves in Fig. 4 show the saturation curves at the temperature T_1, T_2, $T_3\cdots$ $(T_1<T_2<T_3\cdots)$. Deterioration of vacuum seldom occurs if the outgassing of the anode is done carefully in the course of pumping. The measurements are carried out in vacuum of $3\sim5\times10^{-7}$ mm Hg. In order to extinguish activation centers such as F-centers, if they are temporarily formed by the irradiation of the x-ray beam on the cathode, the measurement of the crystal size is carried out last. The following heat treatment of the cathode, which is associated with the next stage of the measurements, will extinguish this disturbance.

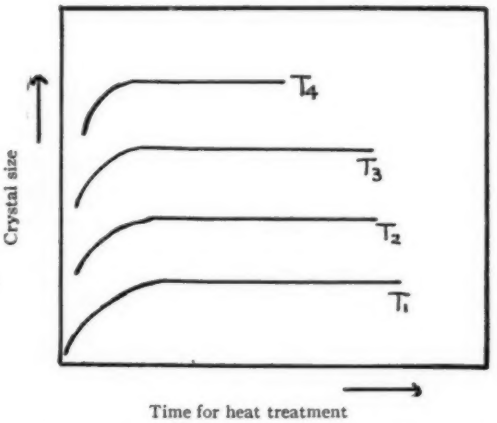


FIG. 4. Schematic crystal growth *vs* the time for heat treatment at various temperatures T_1, T_2, $T_3\cdots$ $(T_1<T_2<T_3\cdots)$.

III. RESULTS

The results of the measurements of the crystal size of an equal molar barium-strontium oxide solid solution are shown in Fig. 5. It is seen from this figure that at a comparatively low temperature, such as 920°C, the time required to reach an equilibrium size is as long as one hour, but at a high temperature, such as 1090°C, an equilibrium size is attained quickly, in about two minutes. The relations between the heating temperature and the equilibrium size are shown in Fig. 6. It is seen from this figure that up to 1000°C the crystal growth is gradual, and above 1050°C the crystals grow very rapidly.

One of our test tubes indicated the separately diffracted beams of the barium and strontium oxides. It indicates a mechanism of decomposition of barium-strontium double carbonate into the oxide, but it can not be concluded from the above that the double carbonate decomposes into barium and strontium oxide and then both oxides combine into double oxide.

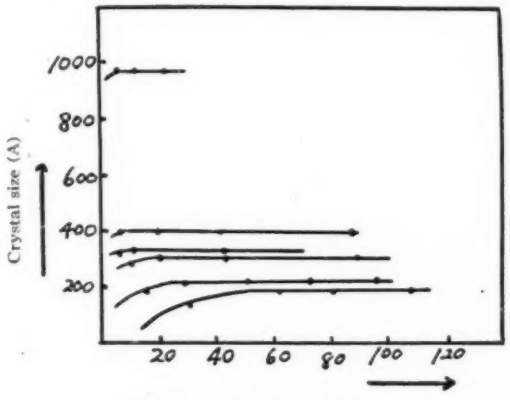


FIG. 5. Experimental crystal growth *vs* the time for heat treatment at various temperatures.

Kawamura[5] had given an excellent discussion of the relation between the crystal size of the oxide and its thermionic activity. He ascertained[1] from his experiments that the heat of evaporation of the equal molar barium-strontium oxide was smaller as much as 1.0 ev than that of barium oxide and ascribed this reason to the difference of crystal sizes, that is, to the seemingly larger size of barium oxide than that of barium-strontium oxide. Using a cyclic process containing solid and gas phases of the oxide, he had concluded that E, activation energy which is required to decompose one BaO into one activation center and half-molecule of oxygen, of the large crystal was 5.4 ev and E of the small crystal is 4.4 ev, and, therefore, the crystal growth caused the decrease of activation centers, as seen from the next formula,

$$N_0(T) = \mathrm{cp}^{-\frac{1}{2}} \exp^{-E(T)/kT} \quad (T = T_1, T_2, T_3 \cdots),$$

[5] H. Kawamura and others, *Electron Emission and Semi-conductor* (in Japanese) 1949, pp. 245, 248.

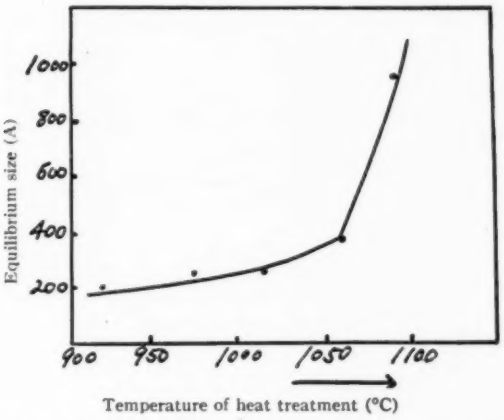


FIG. 6. Equilibrium values of crystal size *vs* temperature of heat treatment of the cathode.

where N_0 is the number of activation centers and p is the partial pressure of oxygen gas.

N_0 at various temperatures for large and small crystals must be determined in order to test Kawamura's conclusion. But in our experiments the equilibrium size is determined by the temperature, and the temperature can be used as a variable instead of the crystal size. If we assume that 4.4 ev is E of 200A crystal size and 5.4 ev is that of 1000A crystal size and moreover E varies linearly with crystal size, then Kawamura's conclusion demands that the curve of $\log N_0$ *vs* $1/T$ is concave downward and has a maximum point showing the existence of the preferential temperature of heat activation, as seen from Fig. 7.

When the temperature of heat treatment of the cathode is raised from 900°C to 1100°C, the activity becomes better and the work functions obtained from the Richardson line are almost the same about 1.6 ev, for the most part of the temperature, range except for the high temperature, where the work functions become smaller than that of the low temperature range, 1.2 ev. If the existence of the frozen-in Frenkel or Schottky type lattice defect is assumed from the comparatively

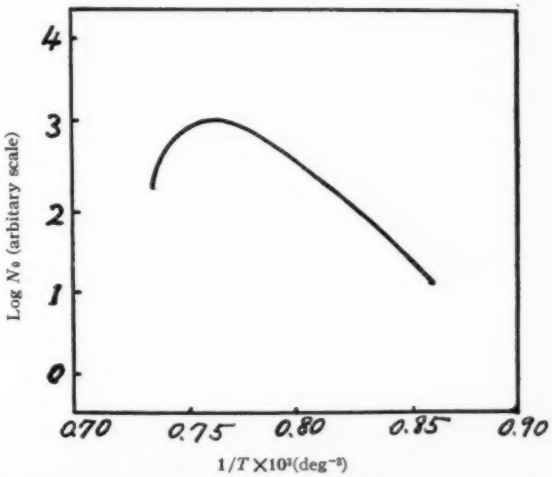


FIG. 7. LogN_0 *vs* $1/T$ (theoretical).

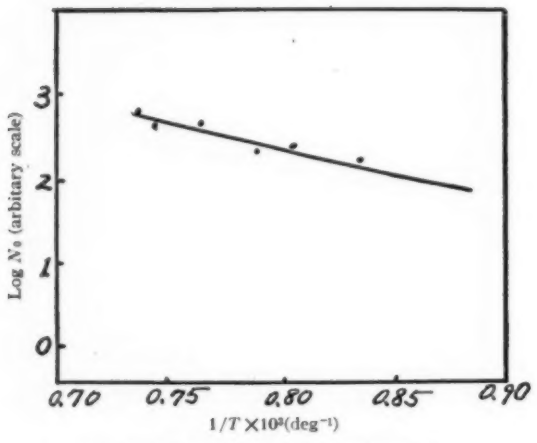


FIG. 8. LogN_0 *vs* $1/T$ (experimental).

large work function, the thermionic current J is

$$J = c' N_0 \exp^{-(x+\epsilon)/kT'},$$

where x is the energy difference between the bottom of the conduction band and vacuum level, ϵ is the energy difference between the impurity energy level and the conduction band, and T' is not the temperature of heat treatment of the cathode (T_1, T_2, $T_3 \cdots$) but the temperature used in the Richardson plot. Therefore, the experimental curve of log$N$ *vs* $1/T$ can be plotted by using N_0, which is calculated from the Richardson constant multiplied by constant c', in the temperature range of the same work functions where constant c' is thought to be the same. The experimental curve shown in Fig. 8 is different from the theoretical one in Fig. 7 and is nearly a straight line. As above mentioned, Kawamura's conclusions are concerned with the case of heat activation, so we checked the existence of activating impurities in the Ni sleeve by a spectroscopic method and recognized impurities of Mg and Si. Therefore, we may conclude that the contradiction between the experimental results and Kawamura's conclusions lies in the presence of activating impurities. Unfortunately, the quantity of impurities is obscure, and we cannot analyze Fig. 8 to reach a conclusion. Nevertheless, it is thought that at lower temperatures impurity activation is dominant and at higher temperatures heat activation is dominant from comparing Fig. 7 and Fig. 8.

It is true that the fact has been ignored in the above discussion that the growth of crystals is caused by the heat treatment in our experiments and by the introduction of water vapor and chlorine gas in Kawamura's experiment, yet the difference in methods of developing crystal growth by no means changes the above results. Also the difference between crystal growth by heat treatment and by the presence of different components of the oxide will not change the above discussion.

When we continue the heating of the cathode, the surface will be richer in SrO than the part near the sleeve, because of the preferential evaporation of barium oxide from the surface. Consequently, the intensity curve of the diffracted x-ray beam is gradually distorted, and the maximum of curve moves towards the position of strontium oxide from that of barium-strontium oxide. The evaporation of the oxide and the variation of composition, therefore, can be calculated as a function of heating temperature, and the depth below the surface from the asymmetrical shape of curve and the maximum, as Eisenstein did. But, unfortunately, the time of heat treatment for a deformed line shape is too long and we did not study it.

IV. CONCLUSION

By constructing a vacuum tube with two mica windows, we could measure directly and repeatedly the crystal growth of one cathode by heat treatment. We found that up to 1050°C the crystal grows slowly, but above, the crystal grows very rapidly to a dimension larger than 1000A. The experimental data gave direct evidence for the conclusion that above 1050°C the crystals grow very rapidly as was inferred by Eisenstein. We also found that in a cathode with the usual sleeve containing an activating impurity, impurity activation is dominant at lower temperatures and heat activation is dominant at higher temperatures.

In a succeeding paper, we shall report the results on barium oxide and barium-strontium-calcium oxide. The author is very grateful to Mr. T. Ichimiya and Mr. T. Shindo for their kind suggestions and discussions throughout this experiment. Also he is indebted to Mr. T. Tanaka and Mr. S. Hirota for their valuable assistance in this work.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Turbulent Diffusion and Erosion

RICHARD W. DAVIES
1200 South Marengo Avenue, Pasadena, California
(Received October 15, 1951)

A new theory of turbulent diffusion is developed so as to describe the dynamics of particulate matter in fluid suspensions. In order that the statistical description be consistent with force equation dynamics of sedimentation, a new diffusion equation is derived from the postulates of a past-future stochastic process.

A steady-state suspension of particles in a turbulent stream is treated like an atomosphere. Dimensional arguments are used to retain the most relevant physical entities in a linear theory. It is shown that the bed load and mid-stream suspension of a muddy stream cannot be considered separately. Some experiments are discussed in this connection. A qualitative theory of sand rippling is explained in terms of the instability of a flat sandy bottom under certain turbulence conditions. A singular perturbation method is employed to obtain an asymptotic expansion of the solution of the equilibrium equation.

Assumptions are discussed and criticized. The advantage of the theory is that all statistical coefficients are averages of familiar dynamical quantities.

INTRODUCTION

THERE are a number of important phenomena in nature about which little more is known than that they are examples of turbulent diffusion. Perhaps the most striking example is erosion and sedimentation. Together they constitute man's most challenging engineering problem. An adequate theory of turbulent diffusion would do much to clarify the peculiarities of this most critical problem.

Part I of this paper introduces a new approach of turbulent diffusion. Part II treats a one-dimensional model of mass transport phenomena independently of Part I. The reader can begin with either section. Those familiar with the etiquette of the mathematical approach to physical problems should begin with Part I. On the other hand, the reader who is more accustomed to the somewhat flexible and arbitrary techniques of engineering sciences and who is not too concerned with formality might be better advised to begin with Part II. Part III attempts to unite the theory developed in Part I with the modeling technique of Part II so as to formulate a real three-dimensional problem in turbulent diffusion.

Most of the research in turbulent diffusion has been connected with a generalization of the equation of heat conduction. The very successful studies of molecular scale diffusion by Einstein,[1] Smoluchowski,[2] and Chapman have without doubt given considerable impetus to research along these lines. Meteorologists, however, have departed from this more classical view in search for a better explanation of their observational data.

The equation of heat conduction (also called the diffusion equation of the Ficksian type) is a limiting form of the difference equation derived from the simplest random walk problem. The most characteristic feature of the random walk is the way in which a particle changes its position. A particle at point P moves to $P+\Delta P$ within some specified time interval Δt say. When it moves it jumps with infinite speed. Its average speed is finite but its true speed is either zero or infinity. G. I. Taylor[3] appears to have been the first to suggest that it is the rapid change in direction of motion and not quick changes in position that characterizes turbulent diffusion. He endeavored to attract the attention of mathematicians to this fact with his well-known publication of 1922. The most significant response was made recently by Goldstein[4] who generalized Taylor's simple model of turbulent diffusion and demonstrated that in the limited case diffusion is described by the telegrapher's equation. Goldstein indicates that his equation is only indirectly related to turbulent diffusion. He further indicates that additional generalization is possible.

In Part I it is shown that by extending the postulates of Chapman and Smoluchowski (as contrasted with the discrete approach of Einstein and Goldstein) on the existence of transition probabilities, another more general diffusion equation can be derived. By neglecting higher order derivatives that have small coefficients the equation reduces to a hyperbolic differential equation.

The steady-state form of the classical diffusion equation is a velocity or momentum equation. In the problem of sedimentation, for example, this states that the velocity of settling particles is balanced by a velocity of diffusion. The steady-state form of the new diffusion equation is an acceleration or force equation. Again in the case of sedimentation this is interpreted as the force of gravity being equated to the gradient of a particle pressure.

PART I

Section 1

Let $V(x, t)$ designate the concentration or number density of particles at the point (x, t). $V(x, t)$ can be some other scalar function as well, such as the fraction of space occupied by particles with some specific prop-

[1] A. Einstein, *Theory of Brownian Movement* (E. P. Dutton and Company, New York, 1905–1906), with notes by R. Fürth.
[2] M. von Smoluchowski, Physik. Z. **17**, 557–585 (1916).
[3] G. I. Taylor, Proc. London Math. Soc. (2) **20**, 196–212 (1921–1922).
[4] S. Goldstein, Quart. J. Mech. and Applied Math. **IV**, Pt. 2, 130–156 (1951).

erty. The space variable x can be regarded as a position vector, but in the ensuing analysis we shall employ only one dimension. Let $\xi=y-x$ denote the distance a particle moves in the time interval τ. We next define $f(y-x, \tau, y, t)$ as the (transition) probability that a particle definitely located at y at time t will go to x at time $t+\tau$. The transition probability is also presumed to satisfy the conditions

$$f(y-x, \tau, y, t)\rightarrow\delta(x-y) \quad \text{as} \quad \tau\rightarrow 0, \tag{1}$$

$$f(y-x, \tau, y, t)\geqslant 0, \tag{2}$$

$$\int f(y-x, \tau, y, t)d(y-x)=1. \tag{3}$$

With the foregoing assumptions, the starting point in the classical diffusion theory of Chapman and Smoluchowski is the integral equation,[5]

$$V(x, t+\tau)=\int f(y-x, \tau, y, t)V(y, t)d(y-x). \tag{4}$$

Section 2

We now depart from the classical theory by defining another transition probability $g(y-x, -\tau, y, t)$ as the probability that a particle definitely located at y at time t came from x at time $t-\tau$. We assume g satisfies the properties (1), (2), (3). It follows that we can write another integral equation:

$$V(x, t-\tau)=\int g(y-x, -\tau, y, t)V(y, t)d(y-x). \tag{5}$$

Combining Eqs. (4) and (5) into a single average equation, then

$$\frac{V(x, t+\tau)+V(x, t-\tau)}{2}$$

$$=\int\frac{f(y-x, \tau, y, t)+g(y-x, -\tau, y, t)}{2}\times V(y, t)d(y-x). \tag{6}$$

If all the functions are analytic in all their variables, they can be expanded into power series, such as

$$V(x, t+\tau)=V(x, t)+V_t(x, t)\tau+V_{tt}(\tau^2/2!)+\cdots,$$

$$V(x, t-\tau)=V(x, t)-V_t(x, t)\tau+V_{tt}(\tau^2/2!)+\cdots,$$

$$f(y-x, \tau, y, t)V(y, t)$$

$$=\sum_{n=0}^{\infty}(1/n!)\{(x-y)^n(\partial/\partial x)^nV(x, t)f(y-x, \tau, x, t)\},$$

$$g(y-x, \tau, y, t)V(y, t)$$

$$=\sum_{n=0}^{\infty}(1/n!)\{(x-y)^n(\partial/\partial x)^nV(x, t)g(y-x, -\tau, x, t)\}.$$

Thus, Eqs. (4) and (5) can be written, respectively, as

$$\frac{\partial V}{\partial t}=\frac{\partial}{\partial x}(a_1V)+\frac{1}{2!}\frac{\partial^2}{\partial x^2}(a_2V)+\frac{1}{3!}\frac{\partial^3}{\partial x^3}(a_3V)+\cdots, \tag{7}$$

$$-\frac{\partial V}{\partial t}=\frac{\partial}{\partial x}(b_1V)+\frac{1}{2!}\frac{d^2}{dx^2}(b_2V)+\frac{1}{3!}\frac{d^3}{dx^3}(b_3V)+\cdots, \tag{8}$$

where

$$a_n=\lim_{\tau\rightarrow 0}\int\frac{(x-y)^n}{\tau}f(y-x, \tau, x, t)d(y-x)$$

and

$$b_n=\lim_{\tau\rightarrow 0}\int\frac{(x-y)^n}{\tau}g(y-x, \tau, x, t)d(y-x)$$

are assumed to exist. Equation (7) is familiar in diffusion theory. Equation (6) can be written as a differential equation also:

$$\frac{1}{2}\frac{\partial^2V}{\partial t^2}=\frac{d}{dx}(\alpha_1V)+\frac{1}{2!}\frac{d^2}{dx^2}(\alpha_2V)+\frac{1}{3!}\frac{d^3}{dx^3}(\alpha_3V)+\cdots, \tag{9}$$

where

$$\alpha_n(x, t)=\lim_{\tau\rightarrow 0}\int\frac{(x-y)^n}{\tau^2}\left(\frac{f+g}{2}\right)d(y-x)$$

exists for all n.

The functional equation,

$$\frac{\partial V}{\partial t}=\lim_{\tau\rightarrow 0}\left\{\frac{1}{\tau}\int\left(\frac{f-g}{2}\right)V(y, t)d(y-x), \right. \tag{10}$$

is equivalent to subtracting Eq. (8) from Eq. (7) if the various probabilities are sufficiently smooth functions.

If $f=g$, for all τ, then $V(x, t+\tau)=V(x, t-\tau)$; hence, the system is stationary in time.

Section 3

Equation (4) can be written as

$$\frac{V(x, t+\tau)+V(x, t-\tau)}{2}+\frac{V(x, t+\tau)-V(x, t-\tau)}{2}$$

$$=\int\left(\frac{f+g}{2}+\frac{f-g}{2}\right)V(y, t)d(y-x). \tag{11}$$

If one assumes that the separation of the integral on the right into two parts is valid, Eq. (11) can be expanded

[5] For derivation, see S. Chapman, Proc. Royal Soc. (London) **119**, 34–54 (1928).

into the partial differential equation

$$1/2V_{tt}+\lim 1/\tau^2\left\{\frac{V(x,t+\tau)-V(x,t-\tau)}{2}-\int\left(\frac{f-g}{2}\right)V(y,t)d(y-x)\right\}$$
$$=\frac{\partial}{\partial x}(\alpha_1 V)+\frac{1}{2!}\frac{\partial^2}{\partial x^2}(\alpha_2 V)+\frac{1}{3!}\frac{\partial^3}{\partial x^3}(\alpha_3 V)+\cdots,$$

provided the limits α_n and the left-hand side of Eq. (12) all exist.

Section 4

We see that a prediction process (4) does not give any immediate information about the underlying diffusion process. The existence or nonexistence of the various limits depends on the type of motion experienced by the discrete particles. For sake of clarity and argument, let us contrast two kinds of motion of discrete particles. One type is commonly called the random walk. The particle moves in jumps. Within some time interval Δt say, the particle jumps a distance d, in a direction prescribed by some probability law. Its average speed is $d/\Delta t$, although the actual movement is with infinite speed. If d and Δt are made to approach zero in such a way that their ratio still exists, the physical process becomes continuous and is called, therefore, "diffusion."

The particle above can travel the same distance, however, by another kind of motion. It can move continuously with a speed $v=d/\Delta t$ and then alter its direction instantaneously. If now d and Δt are made smaller and smaller in such a way that $2d/\Delta t^2$ exists, then another physical process is obtained which is called "diffusion by continuous movements." It is this latter process that approximates nature more closely. The former process is applicable to problems where the scale is of molecular dimensions such as Brownian motion. When the physical dimensions are blown up (as in turbulent diffusion problems) the existence of a differentiable velocity is of primary importance.

It is interesting to consider a few more formal manipulations. Consider the term $1/2(d^2/dx^2)(\alpha_2 V)$ in Eq. (9). Put $\alpha_2=\langle u^2\rangle$, since the second moment is mean square velocity. Differentiating gives

$$1/2(d/d)(\langle u^2\rangle V)=(\langle u^2\rangle V_{xx}/2)+\langle u^2\rangle_x V_x+\langle u^2\rangle_{xx}V.$$

The term we shall find interesting in the above expansion is $\langle u^2\rangle_x V_x$. For example, let us suppose that the Lagrangian derivatives of $\langle u^2\rangle$ and V are zero:

$$\frac{D}{Dt}\langle u^2\rangle=\frac{\partial\langle u^2\rangle}{\partial t}+\left(\frac{\partial\langle u^2\rangle}{\partial x}\right)\left(\frac{\partial x}{\partial t}\right)=0,$$

$$\frac{D}{Dt}V=\frac{\partial V}{\partial t}+\frac{\partial V}{\partial x}\left(\frac{\partial x}{\partial t}\right)=0.$$

Multiplying these expressions together gives

$$(\langle u^2\rangle)_t V_t+(\langle u^2\rangle)_x V_t u+(\langle u^2\rangle)_t V_x u+(\langle u^2\rangle)_x u^2 V_x=0.$$

Taking the average of the equation and rearranging yields

$$-\frac{(\langle u^2\rangle)_t}{\langle u^2\rangle}=(\langle u^2\rangle)_x\langle u\rangle V_t+\langle u^2\rangle_t\langle u\rangle V_t+(\langle u^2\rangle)_x V_x.$$

The term of the left is proportional to the rate of change of energy per unit energy. In the theory of turbulence this is known as

$$\frac{(\langle u_2\rangle)_t}{\langle u^2\rangle}V_t=\lim_{\tau\to}\left\{\frac{1-R(T,t)}{T}\right\}V_t,$$

where $R(T,t)$ is the velocity correlation function. This is a measure of the microscale of turbulence and turns up as one of the coefficients in Goldstein's equation. Thus, under very special circumstances Eq. (9) can be reduced to the telegrapher's equation.

PART II

THE SUSPENSION OF SOLID PARTICLES IN A TUBULENT FLUID

Section 1

Hurst[6] first suggested that sand particles suspended in a turbulent fluid behave like molecules of gas. He assumed the equivalent of a perfect gas law and equated the gradient of particle pressure to the force of gravity. When the turbulent energy density is constant the particle concentration is identical in form to that of the density in an isothermal atmosphere.

We shall likewise treat particle swarms by methods already known in the classical kinetic theory. In some respects the concepts of kinetic theory approach reality more closely for swarms of sand particles than they do for real gases. Sand particles are large enough for their position to be defined without ambiguity. However, they are not large enough to be regarded as obstacles that disturb the fluid more than the fluid disturbs them.

Let the coordinate y be the vertical distance above the bottom of a flow stream of fluid. We shall assume, for the present, that the stream is of finite depth but of infinite width and length. It is also assumed that the structure of the flow field is homogeneous in all directions except the vertical. Lastly, the problem is assumed to possess a statistical steady state.

The rectangular volume enclosed by $(y, y+dy)$ and a unit length in the longitudinal direction will henceforth be referred to as the unit volume. The fraction of this unit volume occupied by sand particles will be referred to as the unknown $V(y)$. The particles are of uniform density ρ_s, but not necessarily of uniform size.

Let the kinetic energy of the sand particles be expressed by $(\rho_s/2)V(y)\langle U_s^2\rangle(y)$. $\langle U_s^2\rangle$ is then the average

[6] Hurst, Proc. Roy. Soc. (London) **A194**, 196–201 (1929).

of square of the particle velocity. The particle pressure in the vertical direction is twice the kinetic energy of the sand particles. This pressure would be felt by a membrane permeable to fluid particles but impervious to sand particles. The gravitational force on the particles of sand is $2\rho_s(\rho_s-\rho_w)gV(y)/(2\rho_s+\rho_w)$. Hence, we can write

$$(d/dy)\{\rho_s V(y)\langle U_s^2\rangle(y)\}+\frac{2\rho_s(\rho_s-\rho_w)gV(y)}{2\rho_s+\rho_w}=0. \quad (12)$$

Section 2

It is natural that the next step in the analysis should be to relate $\langle U_s^2\rangle$ to the energy density of the fluid mixture. Let $\langle U_w^2\rangle$ be the average of the square of the liquid particle velocity. Then, by definition,

$$\rho_s V(y)\langle U_s^2\rangle+\{1-V(y)\}\rho_w\langle U_w^2\rangle = 2\mathcal{T}(y)\equiv 2\times(\text{total energy in unit volume}), \quad (13)$$

$$V(y)\{\rho_s\langle U_s^2\rangle-\rho_w\langle U_w^2\rangle\}=2\mathcal{T}(y)-\rho_w\langle U_w^2\rangle. \quad (14)$$

If now we assume the time average energy is equally distributed by volume in the unit interval, then

$$\rho_s\langle U_s^2\rangle=\rho_w\langle U_w^2\rangle=2\mathcal{T}(y). \quad (15)$$

We can restate this hypothesis in another way. Designate by S any subset (subvolume) of the unit volume. Let $M(s)$ denote the volume (or measure) of this set. The instantaneous kinetic energy in this subset is designated by $E(S, t)$. Then our equidistribution of average energy by volume amounts to saying that there exists a $\tau_0>0$, such that for all $\tau>\tau_0$

$$(1/\tau M(S))\int_{t-\tau/2}^{t+\tau/2}E(S,t')dt'=\mathcal{T}(y).$$

The magnitude of τ_0 is equivalent to the time scale of the turbulence in the system. We suppose that in that length of time the system interchanges energy between sand particles and water particles, so that their averages are the same. Bagnold[7] has succinctly pointed out that a stream has the capacity to pick up and retain a maximum amount of sediment. Once it has arrived at a steady-state suspension the stream can do no more work on the sediment than the sediment can do on the stream liquid. In the steady state a stream can only pick up particles because it is simultaneously depositing them.

Equation (12) can be rewritten,

$$\frac{1}{\rho_s}\frac{d}{dy}(\mathcal{T}V)+\left(\frac{\rho_s-\rho_w}{\rho_w+2\rho_s}\right)gV=0, \quad (16)$$

which has the solution

$$V(y)=\frac{V(y_0)\mathcal{T}(y_0)}{\mathcal{T}(y)}\left\{\exp-\int_{y_0}^{y}\frac{\rho_s(\rho_s-\rho_w)g}{(2\rho_s+\rho_w)\mathcal{T}(x)}dx'\right\}. \quad (17)$$

[7] R. A. Bagnold, Brit. J. Appl. Phys. **2**, 29–34 (1951).

Section 3

The acceleration Eq. (16) is based upon idealizations of the perfect gas law. It is a first-order differential equation and as such so built as to deal primarily with atmospheres of infinite height. It neglects the forces (influence on the pressure) due to the finite volume of the sand particles and to any kind of interaction between the particles. In the main part of the stream these terms are so small that neglecting them appears to be unimportant. However, it is near the boundaries that these forces manifest their importance. Thus, what occurs in that part of the atmosphere not near the boundaries is influenced by the small forces. This is an instance of small causes having more than just small effects.

Since we intend to deal with streams of finite depth, the argument in the previous paragraph leads us to suspect that perhaps we need more than one boundary condition for a proper determination of the atmosphere's concentration. Another boundary condition requires a higher order derivative in the force equation. Let us see if we can add an acceleration term to Eq. (16) by dimensional reasoning.

The simplest modification of Eq. (16) is to try to include the effects of viscosity of the fluid on the suspension atmosphere. The acceleration on a sand particle due to viscosity is proportional to the kinematic viscosity, ν, times a velocity difference. The dimensions of kinematic viscosity multiplied by velocity is that of acceleration times length squared. $V(y)$ is dimensionless, so that we can regard it as an independent dimension. We desire a term that is dimensionally homogeneous with the other two terms of Eq. (16). By being dimensionally homogeneous we mean that the resulting equation is invariant under any changes of scale in the dimensions, length, time, and concentration. We take the velocity difference (C), referred to as

$$C(y)=(\langle U_w^2\rangle)^{\frac{1}{2}}-(\langle U_s^2\rangle)^{\frac{1}{2}}, \quad (18)$$

$$C(y)=[1-(\rho_w/\rho_s)^{\frac{1}{2}}](\mathcal{T}(y)/\rho_w)^{\frac{1}{2}}. \quad (19)$$

We write as a new acceleration equation

$$\nu\alpha\frac{d^2}{dy^2}\left[\left(\frac{\mathcal{T}}{\rho_w}\right)^{\frac{1}{2}}V\right]+\frac{d}{dy}\left(\frac{\mathcal{T}V}{\rho_s}\right)+\frac{(\rho_s-\rho_w)}{\rho_w+2\rho_s}gV=0, \quad (20)$$

where α is now a dimensionless number depending on the shape of the particles as well as on the difference in densities of the liquid and suspended particles. The new term in Eq. (20) is the resultant force of the interaction of the particles via the liquid. This term was deduced without reference to any law of force between particles and as such is open to criticism. It is linear differential, however, and consequently opens the door to further speculation. It contains the important information that the interaction forces only show themselves macroscopically when the particle distribution is non-uniform. It will be seen that in the middle depths of the stream

this force is small and does little to modify the expression given by Eq. (17), but at the boundaries it becomes of the same order as the acceleration due to gravity.

Equation (20) was derived from assumptions of equidistribution of average kinetic energy and invariance with respect to scale. In order that these assumptions be compatible, we must assume the equidistribution principle for scales of turbulence greater than or equal to the diameter (r) of the smallest particles. The energy in the smaller scales may tend to rotate a sand particle. In the main, however, the smaller scales represent degrees of freedom not open to the solid particles, so that in using the equidistribution assumption care must be taken that $\mathcal{T}(y)$ does not include the energy of the largest wave numbers of the turbulence spectrum. Thus, $\mathcal{T}(y)$ should be defined as,

$$\mathcal{T}(y)=\int_{k=0}^{k=1/r} F(k)dk.$$

Section 4

Equation (20) can be modified to include the influence of the finite volume of the sand particles. The simplest generalization of this kind is found by referring to the equation of state of Van der Waals':

$$Pv=\frac{KT}{1-(b/v)}-F, \tag{21}$$

where v is the volume of the vessel, b is four times the volume of the molecules, KT is the kinetic energy of the molecules, and F is a term due to the attractive intermolecular forces. In our problem the volume of the vessel is fixed at unity and b is allowed to vary; thus,

$$b/v=4V(y).$$

If we use the term derived in the last section as the analogy of the gradient of F, then we can write pressure as

$$\frac{P(y)}{\rho_s}=\frac{V\mathcal{T}/\rho_s}{1-4V}+\alpha\nu\frac{d}{dy}(\mathcal{T}/\rho_s)^{\frac{1}{2}}V; \tag{22}$$

hence, the force equation which is now nonlinear becomes,

$$\nu\alpha\frac{d^2}{dy^2}[(\mathcal{T}/\rho_s)^{\frac{1}{2}}V]+\frac{d}{dy}\left\{\frac{\mathcal{T}}{\rho_s}\cdot\frac{V}{1-4V}\right\}+\left(\frac{\rho_s-\rho_w}{2\rho_s+\rho_w}\right)gV=0. \tag{23}$$

Section 5

In the light of the preceding analysis, it is natural to inquire if in the regions of high concentration and low particle energy density there occurs a change of state. The bottom of the stream is closely packed sand so that a change of state has certainly occurred at the point $y=0$. However, we are interested if such a change of state is possible at a point $y>0$. The upper limit to the region in which solidification or liquefaction might occur will be referred to as the critical height. At the critical height $d^2V/dy^2=dV/dy=0$. Substituting these values into Eq. (20), we find the critical height as the root of the following equation:

$$\alpha\nu\frac{d^2}{dy^2}\left(\frac{\mathcal{T}}{\rho_s}\right)^{\frac{1}{2}}+\frac{1}{\rho_s}\frac{d\mathcal{T}}{dy}+\frac{\rho_s-\rho_w}{2\rho_s+\rho_w}g=0. \tag{24}$$

Equation (24) is independent of V because Eq. (17) is linear. When T is a constant, the critical height is $y=0$. Thus a sand atmosphere, with a single energy level, consists of gaseous suspension only. In actual streams, $\mathcal{T}'$ is always greater than zero near $y=0$. Thus, a critical height above $y=0$ only exists when

$$\frac{d^2}{dy^2}\left(\frac{\mathcal{T}}{\rho_w}\right)^{\frac{1}{2}}$$

is negative.

If $\mathcal{T}(y)$ increases according to a law given by

$$\mathcal{T}(y)=(ay+b)^{\rho}, \quad a, b>0, \tag{25}$$

then

$$\frac{d^2}{dy^2}(\mathcal{T}^{\frac{1}{2}})<0, \quad \text{for} \quad 0<\rho<2. \tag{25a}$$

The root of Eq. (23) is obviously not the most accurate value for the critical height, but it is an upper limit to the true value and as such has the merit that the direction of its error is known.

Section 6

In the last paragraph the implication was made that between the critical height and lower boundary $y=0$, a phase might exist in addition to the gaseous one. This can be expressed in the opposite manner, namely, if there are no roots of Eq. (24) $y_0>0$, then only two phases can exist in equilibrium with one another. These are the gas and the solid phases, and they have an interface which we arbitrarily designate as $y=0$. However, if three phases coexist in the steady state, not one but three interfaces have to be accounted for. Thus, when Eq. (24) has a positive root, the system must set up three interfaces or arrange its bottom geometry so as to exclude the third phase.

It is perhaps worthwhile to restate conditions in terms of Gibbs' phase rule,

$$Ph+F=C+n.$$

Ph is the numbers of phases, F is the variants or degrees of freedom, and C is the number of components. In most problems $n=2$, but in our problem the gravitational force on the system increases n to three. If there is only one component, then the sand particles can exist in, at most, three different phases. Since the concentration $V(y)$ is a function of $\mathcal{T}(y)$ and the two boundary conditions, there are three variants. If we specify $\mathcal{T}(y)$ and

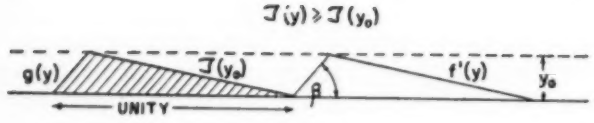


FIG. 1.

two boundaries, then there can be only one phase. If $\mathcal{T}(y)=0$, this phase is a solid. (The case of interest, however, is where the single phase is a gas.) If we insist upon two phases, then one boundary (i.e., one interface) and $\mathcal{T}(y)$ can be specified only. For three phases no boundaries can be specified, or one boundary can be given, but $\mathcal{T}(y)$ cannot be given. This means that for the most general type of bed load we cannot assign boundary conditions. The boundary conditions are determined solely by the intrinsic properties of the sand liquid substance at the energy density $\mathcal{T}(y)$. Since this complicated piece of information is not available, and because it is the role of nature to impose restraints, we shall use an alternative device. We shall propose that the flat boundary $y=0$ is unstable when there is a critical height to $y_0>0$. The solid-gas interface reshapes itself so as to exclude the third phase. In order to do this it must increase the energy density in one locale so as to keep the phase gaseous and diminish the energy density in another region so as to keep the state solid, without changing the total amount of energy in the unit volume. If the fluid mixture builds up mounds of solid sand, then the energy will be packed into the troughs. In the troughs there will be a pick up of solid particles, and on the crest there will be a redeposition of them.

Section 7

It has been suggested that near the lower boundary the sand fluid mixture be considered a heavy viscous fluid with a fairly sharp interface between it and the lighter liquid, which moves across this interface causing rippling. The wavelength of these ripples would be determined by the velocity gradient and surface tension. The latter is a function of $\mathcal{T}(y)$ and $V(y)$.

In our analysis we have assumed a unit length (in order to define a unit volume) which corresponds to the length (or multiple) of a sand ripple, but this length is synonomous to a scaling length of the turbulence $\mathcal{T}(y)$. In order to define $\mathcal{T}(y)$ as an average we had to average over a proper length or its equivalent in the time scale, (an ergodic type of hypothesis). In large streams, such as the Mississippi River or the air flow over the desert, the scale of the turbulence is relatively large (and therefore undamped) so that large sand waves would be expected. In small flumes, where the geometry is very restricting, the length of the largest sand ripples is smaller than the width of the flume.

Section 8

If y_0 is the critical height, the critical energy density $\mathcal{T}(y_0)$ can be found from the energy law, Eq. (25) say. The lower boundary shapes itself so that below y_0 the energy density is $\mathcal{T}(y_0)$. Let $f(y)$ be the shape of the bottom in a given unit interval. Then,

$$\mathcal{T}(y)/[1-f(y)]=\mathcal{T}(y_0). \tag{26}$$

In real flow the shape of ripples resembles the configuration in Fig. 1. The angle β is the angle of repose of the wet sand particles. The function $f(y)$ can be rewritten as the sum of two functions:

$$f(y)=-g(y)+f'(y),$$

where $g(y)\equiv y-y_0/\tan\beta$ and $f'(y)$ is to be determined.

Thus,

$$f'(y)=g(y)+f(y)=\frac{y-y_0}{\tan\beta}+1-\frac{\mathcal{T}(y)}{\mathcal{T}(y_0)}. \tag{27}$$

Section 9

We now attempt to find a solution of Eq. (20). Let $\psi=(\mathcal{T}/\rho_w)^{\frac{1}{2}}V$, $\alpha\nu=\epsilon$, $(\rho_s-\rho_w)/\rho_w=k$, and $(\mathcal{T}/\rho_w)^{\frac{1}{2}}=A$, then Eq. (20) becomes

$$\psi''+(A\psi)''+(k\psi/A)=0. \tag{28}$$

In theory there is no difficulty in finding a solution to Eq. (28) if A is well-behaved. The usual method of finding a solution would be to express the solution in power series about an ordinary point or an isolated singularity of Eq. (28). The solution will converge, but since ϵ is small the convergence is too slow for the solution to be useful. Furthemore, the power series solution would not tell us how the second-order term in Eq. (28) modifies the solution given by Eq. (17).

We shall employ a technique devised by Latta[8] to find an asymptotic expansion of Eq. (28) in which the first term is given by Eq. (17). However, we need two boundary conditions. The first condition will be that used in Eq. (17). Let us assume that somewhere away from the bottom of the flow stream, where measurements are particularly easy to make, $V(y)$ is known. We shall designate this as the point $y=1$. Thus, $V(y)_{y=1}=V(1)$. For the second boundary condition we put $d[V(0))]/dy=0$ at the bottom of the stream. This condition states that at the bottom of the stream the kinetic energy of the particles is reflected. Now consider the expansion:

$$\psi(y)=\psi_0(y)+\psi_1(y)+\cdots\exp\left\{-\int_0^y A(x)dx/\epsilon\right\}$$

$$\times\{H_0(y)+\epsilon H_1(y)+\cdots\}. \tag{29}$$

Substitute Eq. (29) into Eq. (28) and collect terms. Equating each coefficient of ϵ^n and

$$\epsilon^n\exp\left\{-\int_0^y (A/\epsilon)dx\right\}$$

[8] Gordon Latta, "Singular Perturbation Problem," thesis (California Institute of Technology, 1951).

to zero, we find

$$(A\psi_0)'+k/A\psi_0=0, \tag{30a}$$

$$(A\psi_1)'+k/Z\psi_1=\psi_0'', \tag{30b}$$

$$(A\psi_n)'+k/A\psi_n=-\psi_{n-1}'', \tag{30c}$$

and

$$AH_0'-k/AH_0=0, \tag{30d}$$

$$AH_n'-k/AH_n=H_{n-1}''. \tag{30e}$$

Latta has shown that if in Eq. (29) we only write on the first n terms; the remainder is of order (ϵ^{n+1}).

The boundary conditions are now written as

$$\psi(1)=\psi_0(1)+\psi_1(1)+\cdots\exp\left\{-\int_0^1 A\,dx/\epsilon\right\}\times\{H_0(1)+\epsilon H_1(1)+\cdots\}, \tag{31a}$$

$$\left.\frac{d(V)\mathcal{T}}{d_y}\right]_{y=0}=\left.(A\psi)'\right]_{y=0}=0$$

$$=A(0)\{\psi_0'(0)+\psi_1'(0)+\cdots\}$$

$$+A(0)\left\{H_0'(0)+\epsilon H_1(0)\frac{(-A)}{\epsilon}\cdots\right\}$$

$$+A'(0)\{\psi_0(0)+\epsilon\psi_1(0)+\cdots\}$$

$$+A'(0)\{H_0(0)+\epsilon H_1(0)+\cdots\}. \tag{31b}$$

The terms that are independent of ϵ in the above equations give

$$(1)\psi_0(1)+\exp-\int_0^1 A/\epsilon dx H_0(1), \tag{32a}$$

$$0=(A\psi_0)'+(AH_0)'-AH_1. \tag{32b}$$

However, $H_0(y)$ is the solution of the first-order (reduced) equation, and the boundary condition at $y=1$ was assigned to this function as in Eq. (17); hence, $H_0(1)=0$. But this implies $H_0(y)\equiv 0$ from Eq. (30d). We can write then

$$\psi(1)=\psi_0(1)=V(1)\mathcal{T}(1), \tag{33a}$$

$$H_1\,0(=1/A(A\psi_0)'$$

$$=\frac{V(1)\mathcal{T}(1)k}{\{(\mathcal{T}(y_0)\}^{\frac{1}{2}}}\exp\left\{k\int_0^1\left(\frac{1}{\mathcal{T}}\right)dx\right\}. \tag{33b}$$

By substituting the solutions of (30a)–(30e) into summing terms, we can write for an asymptotic expansion of the exact solution

$$V(y)\mathcal{T}(y)=\frac{V(1)\mathcal{T}(1)}{1-\epsilon}\exp\left\{k\int_y^1\frac{dx}{A}\right\}$$

$$\times\left\{1+\frac{\epsilon k}{\mathcal{T}(y_0)^{\frac{1}{2}}}\exp\left\{\int_0^y\left(\frac{2k}{\mathcal{T}}-\frac{\mathcal{T}^{\frac{1}{2}}}{\epsilon}\right)\right\}dx\right\}+0(\epsilon^2),$$

which becomes the solution of the reduced equation when $\epsilon\to 0$.

Section 10

This is as far as we can develop the one-dimensional study. We shall see later how it will help in considering the higher dimensional cases. We have formulated a linearized theory. Its complications justify to some extent the various simplifications we have employed.

Originally it was hoped that existing data could be employed to test the validity of the theory present here. Unfortunately, information about the energy $\mathcal{T}(y)$ is not available. The velocity distribution formulas of Karman and Prandtl are not applicable. It is common practice to postulate linear laws over short ranges. This technique leaves two parameters that can be varied at will. It was the author's experience that with so much leeway almost any data could be fashioned to fit the theory. No purpose is served by such a practice. It seems more reasonable to accept the theory for all its limitations and from the data obtain information about $\mathcal{T}(y)$. From a long-range point of view turbulent energy is the more fundamental problem, and the greatest hopes for measuring it lie in the direction of diffusion.

Rim[9] has run some qualitative experiments at the California Institute which tend to confirm the atmospheric approach to sedimentation. He divided flowing streams of sediment by means of underwater walls running parallel to the direction of flow. The height of the walls were varied. When they were high enough to disconnect both sides completely, rippling on the bottom of the streams was definitely uncorrelated. When the walls were only high enough to disconnect the stream bottoms but not the sediment atmospheres, the ripples were aligned almost perfectly. Thus, rippling cannot be a surface phenomenon only. Details of this and other simple but striking experiments are being prepared by Rim for publication.

PART III

DISCUSSION OF FUTURE PROBLEMS

Section 1

In Part II a linearized force equation was constructed by treating suspended particles as an atmosphere. The natural question at this point is, "How does this theory fit with the classical Ficksian diffusion theory?" It was the endeavor to answer this question that led the author to the development of Eq. (9), Part I.

The usual diffusion equation of sedimentation is written in one dimension as

$$\frac{\partial V}{\partial t}-\eta\frac{\partial V}{\partial x}=\frac{\partial}{\partial x}\left(D\frac{\partial}{\partial x}\right), \tag{34}$$

[9] M. Rim, Hebrew University, Jerusalem; Research Fellow, California Institute of Technology.

where η is the settling velocity and D is a diffusion coefficient. The steady-state equation is given by

$$\eta V+D(\partial V/\partial x)=0. \tag{35}$$

As was stated in the introduction, this is a velocity equation. The acceleration Eq. (20) (Hurst's) has the same type of formal solution as (35) above. Equation (20) should enable us to interpret the limit coefficients of the new diffusion equation in two and three dimensions. For the time being two dimensions will suffice to explain the new features brought out by higher dimensions. The new diffusion equation can be written,

$$V_{tt}=\left(\frac{\rho_s-\rho_w}{2\rho_s+\rho_w}\right)gVy+(\langle U^2\rangle V)_{xx}+(\langle v^2\rangle V)_{yy}$$
$$+2(\langle Uv\rangle V)_{xy}+\frac{\nu}{3!}\{[\langle U^2\rangle+\langle v^2\rangle)]^{\frac{1}{2}}V\}_{xxx},$$
$$+\frac{\nu}{3!}\{[\langle U^2\rangle+\langle v^2\rangle)]^{\frac{1}{2}}V\}_{yyy}+\cdots \text{ etc.}, \tag{36}$$

where

$$\begin{bmatrix}\langle U^2\rangle\langle Uv\rangle\\ \langle vU\rangle\langle v^2\rangle\end{bmatrix}$$

is the energy tensor or stress tensor of the particles. This tensor is of fundamental importance in any fluid dynamical study, and it is not surprising that it should arise here in a very natural way.

Equation (36) is a parabolic partial differential equation. If the third and higher order derivatives are neglected it becomes hyperbolic. Let us see what this means physically. If a source of particles are entered into the turbulent stream at the origin at time $t=0$, there is a small effect on the entire turbulent field immediately afterwards owing to the small particle interaction. The main effect is, however, like a wave traveling out from the origin with a decaying tail following the wave front.

Section 2

It is possible to use the new diffusion to describe the turbulent transport of heat. The turbulent fluid can be regarded as a mixture of hot and cold fluid particles. It is presumed that the cold ones are considerably more numerous than the hot ones. The warmer particles are lighter and therefore subject to an inverse gravitational force. There is also the force of the pressure gradient of the hotter particles. The interaction between the particles can be neglected in comparison to the conduction of heat by molecular processes. The molecular conduction of heat is given by the fourth-order derivatives in the general diffusion equation. The dimensions of the coefficient of the fourth derivative is length to the fourth power over time squared. This is the dimension of kinematic viscosity times the coefficient of molecular heat conduction. The differential equation can be written as

$$T_{tt}=(kT)_y+(\langle U^2\rangle T)_{xx}+(\langle v^2\rangle T)_{yy}+2(\langle Uv\rangle T)_{xy}$$
$$+(\nu D_1T)_{zzzz}+\text{etc.}\cdots. \tag{37}$$

The kinematic viscosity is not necessarily a constant under these conditions, since it varies with temperature and density. However, the problem is kept linear by assuming it to be constant. The molecular conduction of heat is given by the well-known equation:

$$T_t=(D_1T)_{xx}+(D_2T)_{yy}+[D_3T]_{xy}. \tag{38}$$

Substituting Eq. (38) into the fourth-order terms of Eq. (37), it transpires that the diffusion equation takes the form

$$T_{tt}=kT_y+(\langle U^2\rangle T)_{xx}+(\langle V^2\rangle T)_{yy}+2(\langle Uv\rangle T)_{xy}$$
$$+\alpha\nu\{T_{xx}+T_{yy}+2T_{xy}\}t, \tag{39}$$

which is very similar to the equation for the propagation of a sound wave in a viscous atmosphere.[10] This equation is only approximate, but it serves to demonstrate that the molecular scale heat transport has a small damping effect on the turbulent scale transport.

Section 3

In Part II it was seen that determining the proper boundary conditions and practical solution of the simplest one-dimensional problem was not trivial. Solutions to Eqs. (36) and (39) can be found by perturbation methods analogous to the one used in Part II if it is possible to determine solutions of the reduced equations and designate the boundary conditions *a priori*. However, the fulfillment of these necessary conditions is by no means straightforward.

It appears at this time that a more specific mathematical formulation of turbulent diffusion will depend on the individual problem at hand. For example, in certain two-dimensional steady-state problems the diffusion equation can be separated into a system of partial differential equations where the singular curves of the system correspond to the boundaries of the domain. The details constitute the subject matter for a future paper.

[10] H. Lamb, Proc. London Math. Soc. **7**, Series 2, 122 (1909).

Radiation from a Dielectric Wave Guide

ROBERT M. WHITMER
Rensselaer Polytechnic Institute, Troy, New York
(Received February 11, 1952)

The radiation from an infinite dielectric slab fed by a line source has been calculated from a branch-cut integral. The modal efficiency of the device as a wave guide is defined as the ratio of the power propagated in a mode to the sum of the modal and radiated powers. For a range of the parameters such that only one mode can be propagated, the modal efficiency has been plotted numerically. Except for extreme values of the parameters, the efficiency runs around 80 percent to 90 percent.

The slab form is not one likely to be used in practice, but the efficiency of a cylinder would be expected to have about the same value.

1. INTRODUCTION

CONSIDER the two-dimensional dielectric slab of Fig. 1, in which is imbedded a current filament. It has been shown in an earlier paper that the field about the slab may be analyzed into a finite number of modal waves, plus a wave given by an integral around a branch cut.[1] It is the purpose of the present paper to show that the branch-cut integral gives the field radiated from the source, and to find the "modal efficiency" of the slab as a wave guide. The modal efficiency is defined as the ratio of the power transported by the lowest mode to the total power output of the source, when the parameters are such that only the lowest mode can exist.

Short lengths of dielectric guides have engineering applications as antennas. Because of the finite length, standing modal waves are set up. In analysis of these antennas, the standing displacement current waves are treated as the source of radiation, just as with wire antennas we treat the standing waves of conduction current as sources. The type of radiation which is considered in this paper is of a different nature, because its source is the exciting element itself—the filament of conduction current which drives the guide. The space distribution of this radiation is, of course, modified by the presence of the guide.

2. MATHEMATICAL FORMULATION

Much of this section is a restatement of the work of the earlier paper.[1]

The dielectric slab is shown in Fig. 1. It has a thickness $2a$, an index of refraction n (electric permittivity $n^2\epsilon_0$, magnetic permittivity μ_0). It extends to $\pm\infty$ in the y- and z-directions. A filament carrying one ampere of current is at $x=b$, $z=0$ and extends to $y=\pm\infty$.

We see that there will be only a y-component of the electric field. The current-density in the source is

$$J=\delta(x-b)\delta(z),$$

where δ is the Dirac delta-function. The electric field E may be related to a scalar, ψ, by

$$E=E_y=i\omega\mu\psi.$$

Then the differential equation for ψ is[2]

$$(\nabla^2+k^2n^2)\psi=-\delta(x-b)\delta(z). \tag{1}$$

Here $k=2\pi/\lambda$, where λ is the free-space wavelength. A time-dependence factor of $\exp(-i\omega t)$ has been suppressed; with this time-dependence $\exp(ikz)$ represents a wave traveling in the positive z-direction.

As a solution of (1) we have

$$\psi(x,z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}v(x,h)e^{ihz}dh, \tag{2}$$

if we require that v satisfy

$$\left(\frac{d^2}{dx^2}+k^2n^2-h^2\right)v(x,h)=-\delta(x-b) \tag{3}$$

independently of h.

It is necessary that each Fourier component v satisfy the same boundary conditions as ψ itself: v must be continuous everywhere, and dv/dx continuous at $x=\pm a$. If we divide the space into the four regions shown in Fig. 1, the solutions of (3) are[3]

$$\left.\begin{aligned} v_1&=A_1e^{ux/a}\\ v_2&=A_2\cos w\frac{x}{a}+A_3\sin w\frac{x}{a}\\ v_3&=A_4\cos w\frac{x}{a}+A_5\sin w\frac{x}{a}\\ v_4&=A_6e^{-ux/a}\end{aligned}\right\}, \tag{4}$$

where

$$u=a(h^2-k^2)^{\frac{1}{2}} \tag{5}$$

and

$$w=a(k^2n^2-h^2)^{\frac{1}{2}}. \tag{6}$$

The boundary conditions give six equations—one of them inhomogeneous because of the presence of the source—from which the six A's may be determined.

[1] R. M. Whitmer, "Fields in nonmetallic waveguides," Proc. Inst. Radio Engrs. **36**, 1105 (1948). In the present paper this will be referred to as "Fields."

[2] This ψ is $1/(2\pi)$ times the ψ used in Fields, which was incorrectly normalized.

[3] The u of Eqs. (4) is the negative of the λ used in Fields.

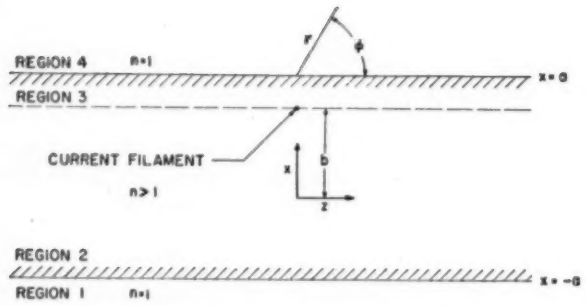


FIG. 1. A dielectric slab, index of refraction $n>1$, extends between $x=\pm a$, and to $\pm\infty$ in the y and z directions. A filament of current at $x=b$, $z=0$ produces an electric field parallel to the surfaces of the slab.

The resulting rather unwieldy expressions for the v's are

$$\left.\begin{aligned}
v_1&=\frac{ae^u}{2}\left(\frac{1}{D_2}\cos w\frac{b}{a}+\frac{1}{D_1}\sin w\frac{b}{a}\right)e^{ux/a}\\
v_2&=\frac{a}{2w}\left(\frac{1}{D_2}\cos w\frac{b}{a}+\frac{1}{D_1}\sin w\frac{b}{a}\right)\\
&\qquad\times\left(D_1\cos w\frac{x}{a}-D_2\sin w\frac{x}{a}\right)\\
v_3&=\frac{a}{2w}\left(\frac{1}{D_2}\cos w\frac{b}{a}-\frac{1}{D_1}\sin w\frac{b}{a}\right)\\
&\qquad\times\left(D_1\cos w\frac{x}{a}+D_2\sin w\frac{x}{a}\right)\\
v_4&=\frac{ae^u}{2}\left(\frac{1}{D_2}\cos w\frac{b}{a}-\frac{1}{D_1}\sin w\frac{b}{a}\right)e^{-ux/a}
\end{aligned}\right\},\qquad(7)$$

where

$$D_1=w\sin w\left[\cot w+\frac{(d^2-w^2)^{\frac{1}{2}}}{w}\right]\qquad(8)$$

$$D_2=w\cos w\left[\tan w-\frac{(d^2-w^2)^{\frac{1}{2}}}{w}\right],\qquad(9)$$

and

$$d=ka(n^2-1)^{\frac{1}{2}}.\qquad(10)$$

3. INTEGRATION; SINGULARITIES

Equation (2) is to be evaluated as a contour integral in the complex h-plane. There are a finite number of singularities, at the roots of D_1 and D_2. These occur on the real h-axis, in pairs symmetrical about the origin. They are given by

$$D_2=0,\quad\text{or}\quad\tan w=\frac{1}{w}(d^2-w^2)^{\frac{1}{2}}\qquad(11)$$

and

$$D_1=0,\quad\text{or}\quad-\cot w=\frac{1}{w}(d^2-w^2)^{\frac{1}{2}}.\qquad(12)$$

Plots of each side of (11) and (12), for some value of the parameter d, are shown graphically in Fig. 2. For a root, w, we calculate the corresponding value of h from Eq. (6).

The contour of integration is shown in Fig. 3. It is closed in the upper half-plane for $z>0$, in the lower half-plane for $z<0$. The residue at each pole gives a wave having the form of

$$f(x)e^{ihz},$$

with h real. This we call a *modal-type* wave. From Fig. 2 it is clear that there is always at least one root of (11), no matter how small the parameter d. To this extent the lowest mode behaves like a principal mode.

Further examination of Fig. 2 shows that only one mode will exist if $d<\pi/2$; throughout the rest of this paper we shall limit ourselves to this situation.

Evaluation of the residues gives, for $z>0$,

$$\left.\begin{aligned}
\psi_1&=-\frac{i}{2}\frac{uw^2}{had^2(u+1)}\frac{\cos wb/a}{\cos w}\exp[u(1+x/a)+ihz]\\
\psi_2=\psi_3&=-\frac{i}{2}\frac{u}{ha(u+1)}\cos w\frac{b}{a}\cos w\frac{x}{a}e^{ihz}\\
\psi_4&=-\frac{i}{2}\frac{uw^2}{had^2(u+1)}\frac{\cos wb/a}{\cos w}\exp[u(1-x/a)+ihz].
\end{aligned}\right\}\qquad(13)$$

4. INTEGRATION; BRANCH CUT

Each of the v's of (4) has a branch point at $h=\pm k$. These arise from the branch points of u [see Eq. (5)]. Although w [Eq. (6)] has branch points at $h=\pm kn$, the v's are even in w, so the integrands do *not* have branch points there. Application of the radiation con-

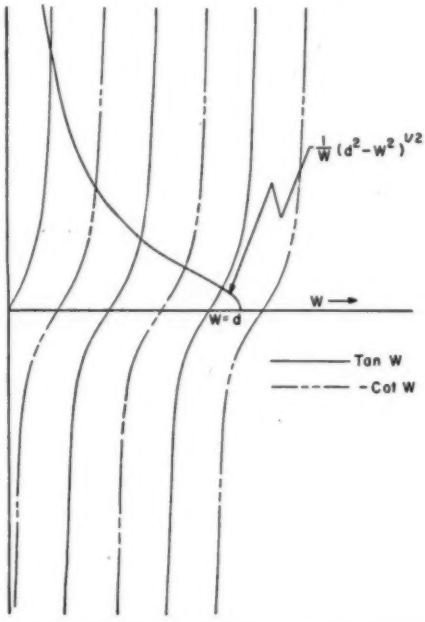


FIG. 2. Curves illustrating Eqs. (11) and (12).

dition tells us that the contour goes above $h=-k$, below $h=+k$. Convergence of the integrals requires that the branch cuts be at least asymptotic to the imaginary axes. It is convenient to make them follow the real axes to the origin, then go out along the imaginary axes.

For $z>0$, the radiated wave in Region 4 is

$$\psi_{4r}=\int_{C_1} f(h)\exp[u(1-x/a)+ihz]dh, \qquad (14)$$

where

$$f(h)=-\frac{a}{4\pi}\left(\frac{1}{D_2}\cos w\frac{b}{a}-\frac{1}{D_1}\sin w\frac{b}{a}\right). \qquad (15)$$

Here C_1 is the contour around the branch cut, $i\infty$ to k and back to $i\infty$. In the D's, and in the exponent in (14), the radiation condition says that u is a positive imaginary on the upper and right side of the cut, a negative imaginary on the opposite side.

The exponent of e in (14) is

$$\pm u(1-x/a)+ihz=\pm i(k^2-h^2)^{\frac{1}{2}}(x-a)+ihz$$

with the alternative signs applying to opposite sides of the branch cut. Let

$$h/k=\cos\zeta; \qquad (16)$$

then

$$(k^2-h^2)^{\frac{1}{2}}=k\sin\zeta \qquad (17)$$

and

$$dh=-k\sin\zeta d\zeta. \qquad (18)$$

Also, from the geometry depicted in Fig. 1,

$$x-a=r\sin\phi \quad \text{and} \quad z=r\cos\phi.$$

Combining these changes of notation,

$$\pm u(1-x/a)+ihz=ikr\cos(\zeta\mp\phi). \qquad (19)$$

When $h=k$, $\zeta=0$; when $h=i\infty$, $\zeta=\pm(\pi/2-i\infty)$. We choose to use the upper sign. Then,

$$\psi_{4r}=-k\int_{C_2} f_-(\zeta)\exp[ikr\cos(\zeta-\phi)]\sin\zeta d\zeta$$

$$-k\int_{-C_2} f_+(\zeta)\exp[ikr\cos(\zeta+\phi)]\sin\zeta d\zeta. \qquad (20)$$

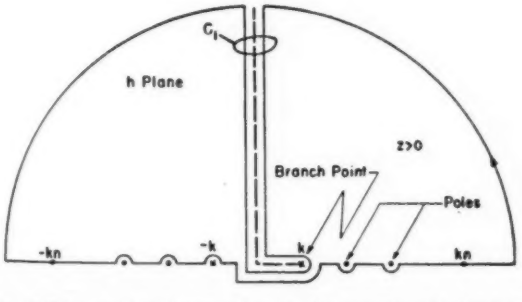


FIG. 3. The path of integration in the complex h-plane. There are poles along the real axis between k and kn. There are branch points at $\pm k$.

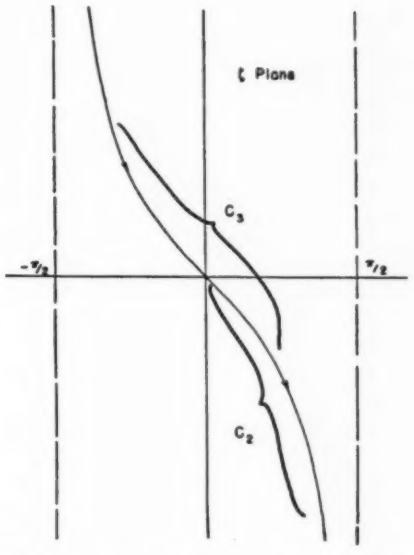


FIG. 4. Contours of integration in the ζ-plane.

The path C_2 is shown in Fig. 4. In $f_\mp$, u and φ have the relative signs given by (19). The symmetry of the integrand is such that the integral over $-C_2$ may be replaced by its image in the upper half-plane:

$$\psi_{4r}=-k\int_{C_3} f(\zeta)\exp[ikr\cos(\zeta-\phi)]\sin\zeta d\zeta, \qquad (21)$$

where the path C_3 is shown in Fig. 4. In (21),

$$f(\zeta)=\begin{cases} f_+(-\zeta) \text{ in the second quadrant} \\ f_-(+\zeta) \text{ in the fourth quadrant.}\end{cases}$$

Finally, let

$$\eta=\zeta-\phi; \quad \zeta=\eta+\phi.$$

Then

$$\psi_{4r}=-k\int_{C_4} f(\eta+\phi)\exp(ikr\cos\eta)$$

$$\times(\sin\eta\cos\phi+\cos\eta\sin\phi)d\eta. \qquad (22)$$

The path C_4 is shown in Fig. 5.

No general integral of (22) has been found. However, there are no singularities of f in the significant region of the η-plane; we are free to cross the real axis wherever we please. If we cross at $\eta=0$, the exponent $ikr\cos\eta$ has a negative real part except at the origin. Then for $kr\gg1$ the integrand is negligible except at $\eta=0$ (there is a saddle-point at the origin) and

$$\psi_{4r}=-kf(\eta=0)\int_{C_4}\exp(ikr\cos\eta)$$

$$\times(\sin\eta\cos\phi+\cos\eta\sin\phi)d\eta. \qquad (23)$$

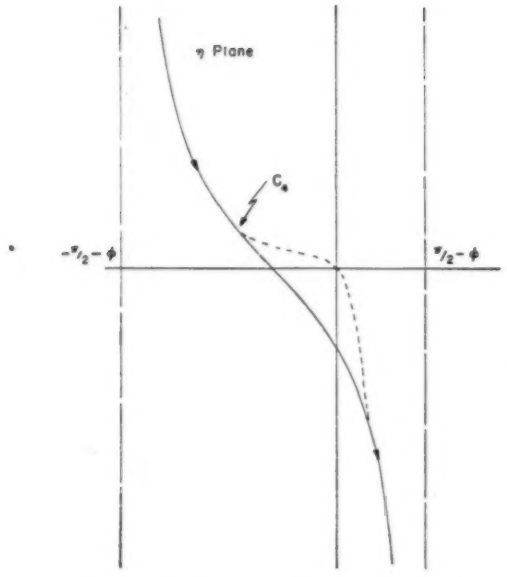


FIG. 5. Contours of integration in the η-plane.

This can be integrated to give[4]

$$\psi_{4r}=-kf\frac{\pi}{2}\{\cos\phi[H_1^{(1)}(kr)+H_{-1}^{(1)}(kr)]+i\sin\phi[H_1^{(1)}(kr)-H_{-1}^{(1)}(kr)]\}$$
$$=-i\pi kfH_1^{(1)}(kr)\sin\phi. \tag{24}$$

5. POWER CALCULATIONS

Differentiation of the modal ψ's of (13) gives the magnetic fields, and calculation of the modal power is then straightforward. However, in order to obtain numerical values it is necessary to have solutions of the eigenvalue equation (11). To obtain these for a set of arbitrary values of the parameter d would be extremely laborious. Fortunately the modal power can be expressed entirely in terms of the internal transverse propagation factor w, with d not appearing explicitly. Also, direct substitution shows that—at least for $d<\pi/2$—the simple relation

$$d=w/\cos w \tag{25}$$

is a solution of (11). The procedure adopted for numerical calculation was to assign arbitrary values to w and compute numerical values of the modal power and of the corresponding values of d. These values of d were then used in the numerical computations of the radiated power.

The radiated power in Region 4 is calculated from the wave function of (24); the radiated power in Region 1 is obtained from that by changing the sign of h, the distance of the source from the z-axis. To the approximation $kr\gg1$, the modal power flow inside the guide is negligible.

[4] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press, Inc., New York, 1949), pp. 86–87.

The expression for the modal power is given by

$$\frac{P_m}{\omega\mu_0}=\frac{s\cos^2wb/a}{4w\left[\dfrac{n^2}{(n^2-1)\cos^2w}-1\right]^{\frac{1}{2}}}\cdot\frac{w^2e^s+s+\sin^2w}{(s+1)^2}. \tag{26}$$

Here ω is the angular frequency of the field, and

$$s=w\tan w. \tag{27}$$

The corresponding expression for the radiated power is

$$\frac{P_r}{\omega\mu_0}=\frac{1}{4\pi}\int_0^{\pi/2}\frac{p+\cos^2q+\left(\sin^2q-\dfrac{b}{a}\right)(2\sin^2q-1)}{(1+p)p+\sin^2q\cos^2q}pd\phi; \tag{28}$$

here

$$p=\frac{\sin^2\phi}{n^2-1}\quad\text{and}\quad q=d\left[\frac{n^2-\cos^2\phi}{n^2-1}\right]^{\frac{1}{2}}. \tag{29), (30}$$

The integral of (28) was evaluated by numerical methods.

The resulting modal efficiencies, $P_m/(P_m+P_r)$, are plotted in Fig. 6.

6. CONCLUSIONS

The slab has been analyzed because its geometrical simplicity makes the mathematics easier to handle than would be the case for other forms—a cylinder, for example. One would expect the modal efficiencies of a cylinder to have roughly the range of values we have

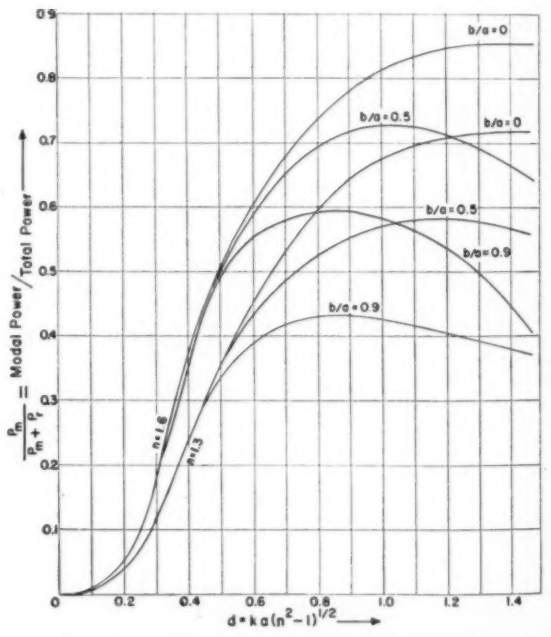


FIG. 6. The modal efficiency of the dielectric guide, showing its dependence upon the parameters of the problem.

found here for corresponding values of the parameters. The power radiated directly from the line source in the slab is small—around 10 percent of the total input. This figure could certainly be reduced considerably by the use of some sort of focusing source.

A dielectric guide made of a flexible plastic may be used as a microwave patch-cord, for connecting different pieces of apparatus. The figures obtained in this paper tell us that such a device may be made to have a high efficiency. (For the patch-cord, "efficiency," and "modal efficiency" mean the same thing.)

We may also apply these considerations to the fields radiated from a dielectric rod used as an antenna. Our results make it plausible that the *directly* radiated field is small compared with that arising from the modal standing waves. Finally, we expect the actual radiation pattern to differ but slightly from that calculated from the modal standing waves alone.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Dimensional Changes Normal to the Direction of Diffusion*

R. W. Balluffi and B. H. Alexander†
Atomic Energy Division, Sylvania Electric Products, Inc., Bayside, New York
(Received April 30, 1952)

Silver from the vapor phase was diffused into thin gold wires and foils at 940°C. Measurements were made of the resulting dimensional changes which occurred normal and parallel to the direction of diffusion. The linear percentage expansion normal to the diffusion direction was always found to be only slightly less than the corresponding expansion parallel to the diffusion stream. The results are best explained by the creation of new lattice sites at dislocations and support both the vacancy and interstitial mechanisms of diffusion in substitutional solutions.

INTRODUCTION

BY the study of marker movements it has been established that the two components in a binary metallic system diffuse at different rates.[1,2] As a result of this unequal flow, one side of a diffusion couple gains atoms at the expense of the other side. This is shown schematically in Fig. 1, where it is indicated that the left-hand or A side of the couple is enlarged in the X direction. The B side of the diffusion couple loses atoms and this may result in a shrinkage as indicated in Fig. 1 or in the development of porosity[3] as shown in the microphotograph of a Cu α-brass diffusion couple in Fig. 2.

The boundary separating the A and B sides is always the moving (originally welded) interface which would be defined by inert Kirkendall markers placed at the original weld. The question naturally arises whether the dimensional changes schematically indicated in Fig. 1 are entirely limited to the X direction (the direction in which the concentration gradient exists), or whether there may also be a change in dimensions in the Y direction normal to the direction of diffusion. DaSilva and Mehl[1] have discussed this question and on the basis of one experiment they conclude that the dimensional changes are probably unidirectional—in the direction of the diffusion stream; they recommend that the problem be studied further.

It is possible that dimensional changes not in the direction of diffusion are restricted in the ordinary sandwich type of diffusion couple (Figs. 1 and 2) by the large mass of metal on either side of the diffusion interface, and that a different specimen geometry would accentuate the effect and allow more accurate measurements. In the experiments described below we have made measurements of dimensional changes normal and parallel to the direction of diffusion in small wires and thin foils of gold into which silver was diffused from the vapor.

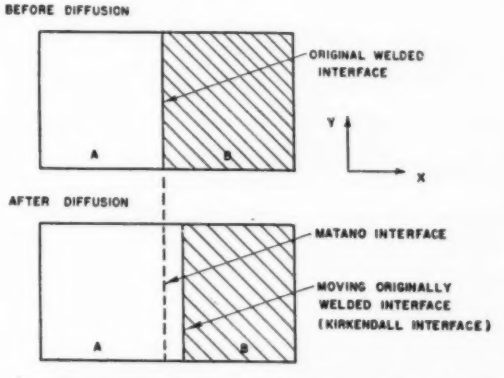


FIG. 1. Schematic diagram of the dimensional changes in a diffusion couple if no porosity is generated and if the changes are restricted completely to the "X" direction.

* This work was sponsored by the AEC and was done under Contract AT-30-1 GEN-367.
† Now at Boston Electronics Division, Sylvania Electric Products Inc., Boston, Massachusetts.
[1] L. C. C. daSilva and R. F. Mehl, J. Metals **191**, 155 (1951).
[2] A. Smigelskas and E. Kirkendall, Am. Inst. Mining and Met. Engrs. **171**, 130 (1947).
[3] R. W. Balluffi and B. H. Alexander, submitted for publication in this journal, 1952.

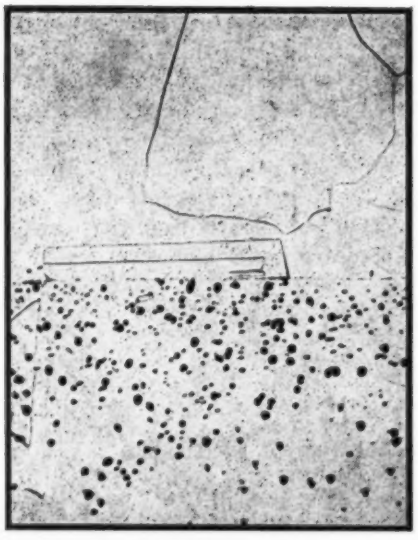


FIG. 2. Porosity developed in the alpha-brass side of a copper/alpha-brass couple after diffusion for 88 hr at 880°C in hydrogen. 50×

EXPERIMENTAL

Gold sheets and wires were silverized by heating them in close proximity to a large reservoir of silver in a continuously evacuated vacuum furnace. Particular care was taken to maintain both the gold specimens and the silver reservoir at the same temperature to eliminate possible condensation of a silver layer upon the gold surface. This method transferred silver atoms from the vapor phase to the gold surfaces at a sufficient rate to maintain the concentration there at approximately 90 percent Ag. The vapor pressure of silver is approximately 5×10^3 times that of gold at the experimental temperatures used and negligible evaporation of gold was found to occur. The silver and gold used were 99.99 percent pure.

TABLE I(a). Expansion of thin polycrystalline gold sheets normal to diffusion direction after heating 2 hours at 950°C in silver vapor.

Sheet no.	Original length (l_0) (cm)	Original width (W_0) (cm)	l_0/W_0	Sheet thickness (inches)	$\Delta l/l_0$ ×(100)	$\Delta W/W_0$ ×(100)
1	1.09	1.04	1.0	0.001	14	14
2	1.09	0.52	2.1	0.001	15	16
3	1.19	0.19	6.3	0.001	16	16

TABLE I(b). Dimensional changes of single and polycrystalline gold wires by heating in silver vapor.

Wire no.	Original diameter (inches)	Temp. (°C)	Time	$\Delta l/l_0$(100) Normal to diffusion	
				Single	Polycrystalline
1	0.0050	940	5 min	···	3.3
2	0.0050	940	15 min	···	9.3
3	0.0065	940	2 hr	5.3	···
4	0.0140	940	3 hr	4.1	7.6
5	0.0140	930	4 hr	4.0	7.7

A. Gold Sheets

Polycrystalline gold sheets 0.001 in. thick and of different length-to-width ratios were heated in silver vapor for two hours at 950°C, and the resulting dimensional changes normal to the diffusion direction are shown in Table I(a). In all cases considerable expansion normal to the diffusion gradient was observed. The percentage expansions in the directions normal to diffusion averaged about 15 percent and appeared to be independent of the geometry of the sheets.

B. Gold Wires

Silver from the vapor phase was diffused into both polycrystalline and single-crystal gold wires 4 cm long and ranging from 0.005 in. to 0.014 in. in diameter, and the resulting percentage linear expansions along the wire axes normal to the diffusion direction are given in Table 1(b).

FIG. 3. Over-etched silver-rich region next to surface of a gold polycrystalline wire, originally 0.014 in. diameter, silverized 3 hr at 940°C. 350×

Polycrystalline wires were found to expand approximately twice as much as single crystal wires in the axial (normal) direction after identical silverization treatments. The average grain size of these polycrystalline wires remained at about one-third of the wire diameter during the diffusion run. Further observations indicated that the greater normal expansion of polycrystalline wires was not due to preferential grain boundary diffusion. Figure 3 shows a typical section next to the surface of a silverized 0.014-in. diameter wire which was over-etched in the silver-rich region up to a certain critical composition. The composition contour shows no preferential penetration into the grain boundaries. In addition, the single crystal and polycrystalline specimens were found to gain the same amount of Ag, showing again that no preferential penetration occurred at the grain boundaries.

Progressive measurements were made of the dimensional changes occurring during the silverization of polycrystalline gold wires originally 0.005 in., 0.014 in., and 0.028 in. in diameter. The percentage length increase of the three wires is shown as a function of the percentage weight gain of silver in Fig. 4. The same

percentage silver penetration in each wire produced about the same percentage normal expansion. Expansion began as soon as silver diffused into the wire surface and therefore long before the diffusion zone reached the center of the wire.

Metallographic examination of the wire cross sections showed that the position of the moving (Kirkendall) interface, which was located at the original surface of the pure gold wire before diffusion, could be detected by careful polishing and etching procedures. This interface was always delineated by a small row of holes and small foreign particles and a discontinuity in grain size (Fig. 5). In a vapor-solid diffusion couple of this type the surface of the pure gold wire before diffusion is termed the original interface, and any foreign particles initially attached to this surface act as moving inert Kirkendall markers during subsequent diffusion and delineate the position of the moving Kirkendall interface. If gold and silver diffuse at the same rate, the markers would remain at the position of

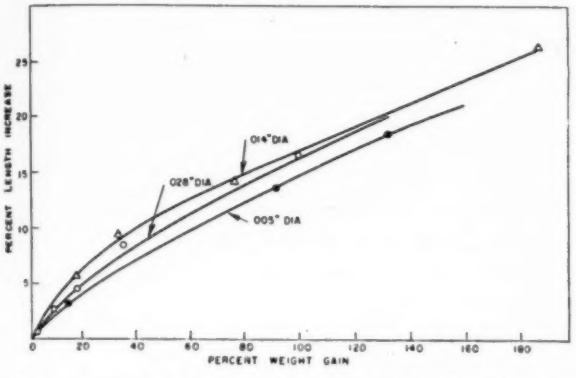


FIG. 4. Percent length increase *vs* percent weight gain of gold polycrystalline wires, originally 0.005 in., 0.014 in., and 0.028 in. diam, as a result of silverization at 940°C.

the original interface. However, if only silver diffuses and gold remains stationary (in a diffusion sense) the wire would expand and the markers would appear on the surface of the silverized wire. As shown in Fig. 5 the Kirkendall interface appears between the position of the original interface (shown dotted) and the surface of the silverized wire. This indicates that silver diffuses more rapidly than gold. This conclusion agrees with the results of daSilva and Mehl,[1] who found that silver diffuses more rapidly than gold on the basis of marker movements in sandwich type Kirkendall specimens. The expansion of the original gold wire in the direction parallel to diffusion is measured by the movement of the Kirkendall interface. The expansion of the wire normal to diffusion is, of course, measured by the longitudinal expansion of the wire minus small corrections for expansion as a result of diffusion into the ends. Data for the percent linear expansion in the two directions of the three wires after diffusion with silver are given in Table II.

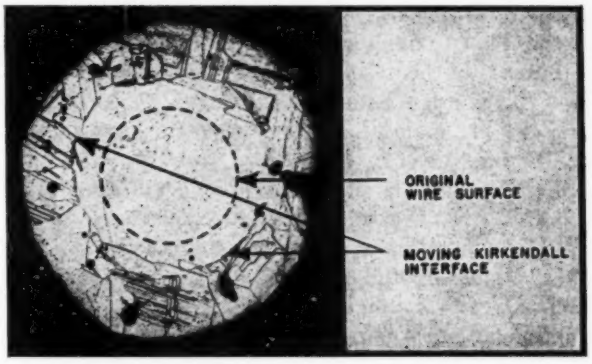


FIG. 5. Etched cross section of gold polycrystalline wire, originally 0.014 in. diameter, silverized 85 hr at 940°C. 100×

The expansions in the two directions are of the same order of magnitude. However, the parallel percentage dimensional changes tend to be generally larger than the normal changes. Unfortunately, the Kirkendall interface was not detectable in the 0.005-in. wire silverized at 940°C for 60 hr, but the close agreement of the normal and parallel expansions in the 0.005-in. wire silverized at 940°C for 85 hr predicts that the expansion due to inward unidirectional diffusion may be isotropic in sufficiently thin specimens.

DISCUSSION

The observation that silver diffuses into the gold wires more rapidly than gold diffuses out, thereby causing the wire to expand in all directions, may be explained by either vacancy or interstitial-diffusion mechanisms. If silver atoms have a higher probability of changing place with vacancies than do gold atoms, an overall vacancy current will flow from the inside of the wire to the outside. This outward flow of vacancies will be compensated for by an equal flow of silver atoms moving inward through it. The generation of vacancies within the wire causes the observed increase in volume. Alternatively, the silver atoms may diffuse into the wire as interstitial atoms which eventually assume substitutional positions, thereby producing the observed ex-

TABLE II. Expansion of Au wires normal and parallel to the diffusion direction as a result of inward diffusion of Ag from the vapor.

Original wire diam (inches)	Diffusion anneal	(100) $\Delta l/l_0$§ (Normal to diffusion)	(100) $\Delta S/S_0$‖ (Parallel to diffusion)
0.005	85 hr at 940°C	35	36
0.014	85 hr at 940°C	24	47
0.028	85 hr at 940°C	21	33
0.005	60 hr at 940°C	...	...
0.014	60 hr at 940°C	30	44
0.028	60 hr at 940°C	20	33

§ l is the wire length corrected for slight expansion due to diffusion into the ends.
‖ S is the diameter of the moving cylindrical Kirkendall interface (see Fig. 5).

pansion. Seitz[4] has shown that Darken's[5] equations, relating the two different diffusion coefficients in binary systems with the resulting Kirkendall marker movement, follow if it is assumed that the vacancies or interstitialcies are maintained everywhere in local thermal equilibrium.

In either event large numbers of vacancies or interstitialcies must be generated and destroyed in the diffusion zone. The most plausible sources and sinks for imperfections of these types are dislocations with Taylor-Orowan character, and the generation and destruction of such imperfections have been discussed in some detail elsewhere.[6–8] The destruction or generation of vacancies (alternatively the generation or destruction of interstitialcies) causes the dislocation line to move normal to its slip surface and this motion has been termed "positive or negative climb[7]." Such movement causes a volume change per unit area swept out by the dislocation line equal to the magnitude of the Burgers vector.

Ideally, if the gold being penetrated by silver could be maintained under uniform triaxial stress during diffusion, and if the dislocations energies were independent of orientation, the local expansion in the diffusion zone due to inward diffusion of silver should be isotropic. Under these conditions the energy necessary for the creation of new lattice sites at dislocations would be independent of orientation and equal numbers of sites would be created at dislocations of all orientations. However, any such expansion caused by penetration of silver in the diffusion zone adjacent to the surface of the silverized wires immediately establishes triaxial tensile stresses in the nondiffused core and mainly compressive stresses near the surface. The stress pattern is analogous to that present in a metallic cylinder after local surface heating. In the expanded volume undergoing diffusion adjacent to the surface, the longitudinal and circumferential stresses are compressive while the radial stress is tensile. Therefore, less energy is required to generate vacancies (or destroy interstitial atoms) at dislocations so oriented to produce a radial expansion. This may account for the radial expansion being somewhat larger than the longitudinal expansion.

The solution of the elastic problem of a non-uniformly heated cylinder indicates that the longitudinal tensile stress acting in the nondiffused core is considerably higher than the other tensile stresses present there. We may, therefore, expect the observed axial lengthening by plastic flow which begins before diffusion reaches the center of the wires. In conventional diffusion specimens where the diffusion zone is small compared to the specimen size, correspondingly high stresses are not built up throughout the bulk of the specimen, and therefore little plastic flow occurs there. Most of the expansion due to diffusion is restricted to the diffusion direction.

The processes by which plastic flow takes place at these temperatures are not completely established. Alexander, Dawson and Kling[9] have shown that deformation occurs by viscous flow at low stress levels near the melting point, and Buttner, Funk, and Udin[10] indicate that at high stress levels both viscous flow and slip occur. Rough calculations show that the type of viscous diffusional flow observed by Alexander, Dawson, and Kling could not produce as high a strain rate as was observed during the present work. It seems likely, therefore, that both of these plastic flow mechanisms operate in the wires during the diffusion anneal. The greater lengthening of the polycrystalline wire is explained on this basis since it is predicted that the flow due to diffusional viscosity would be much greater in polycrystals than in single crystals.[11,12]

ACKNOWLEDGMENT

The writers wish to acknowledge the help of Mr. A. Timper of the Sylvania Metallurgical Laboratories in carrying out most of the experimental portion of this work.

[4] F. Seitz, Acta Cryst. **3**, 355 (1950).
[5] L. S. Darken, Am. Inst. Mining and Met. Engrs. **175**, 184 (1948).
[6] J. Bardeen and C. Herring, Trans. ASM **43A**, 87 (1950).
[7] F. C. Frank, Phil. Mag. **42**, 809 (1951).
[8] N. F. Mott, Proc. Phys. Soc. (London) **64**, 729 (1951).
[9] Alexander, Dawson, and Kling, J. Appl. Phys. **22**, 939 (1951).
[10] Buttner, Funk, and Udin, J. Metals **4**, 401 (1952).
[11] C. Herring, J. Appl. Phys. **21**, 437 (1950).
[12] Work done by A. P. Greenough; reported in Technical Report ONRL 10–52, Office of Naval Research, London, February 18, 1952, p. 7.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

On the Metallic Shadow-Casting Using a Nozzle System

TADATOSI HIBI
Research Institute for Scientific Measurements, Tōhoku University, Sendai, Japan
(Received March 10, 1952)

A new device for metallic shadow-casting by using the nozzle system is proposed, and an outline of the apparatus is given. The present device has a shutter, a nozzle, and a cover plate between specimen and filament in addition to the usual apparatus. The dimensions of the nozzle were chosen from the maximum of 3-mm diam, 1.5-mm diam, and 22 mm to the minimum 1.5-mm diam, 0.7-mm diam, and 38 mm as diameter at the entrance, diameter at the exit, and the length, respectively. It was found that this method was not only conspicuously effective in obtaining much more accurate shadows than that obtained by the usual method with the same shadowing material, but also brought into possible use metals, as shadowing materials, which could not be used for such a purpose by the usual method. Shadowing qualities of many kinds of evaporated metals are compared and a new rule for the order of aggregation of deposited metallic thin films is also proposed.

I. INTRODUCTION

As a specimen technique of electron microscope, the metallic shadow-casting method was established by R. C. Williams and R. W. G. Wyckoff[1] and was developed by these investigators and R. C. Backus.[2] This method has been in universal use as a specimen technique of electron microscopy. The shadowing should not only increase image contrast but also depict the forms and the sizes of specimens three-dimensionally and subsidiarily raise the resolving power of the electron microscope. It is necessary, therefore, to cast a shadow as accurately as possible, but hitherto the shadowing method has not been sufficient for this purpose and the shadowing materials have been limited to the elements U, Au, Cr. Accordingly, I made an improvement here on the usual shadowing method and designed the apparatus using the nozzle system with a shutter. By using this apparatus the shadowing quality of metals hitherto used can be improved pronouncedly and many metals which hitherto could not be used as the shadowing material are brought into possible use for this purpose.

II. PRELIMINARY EXPERIMENT OF SHADOW-CASTING

In order to clarify the effect of the preheating or of vacuum degree for metallic shadow-casting, as a preliminary test, W-shadowing was carried out on various preheating conditions, evaporating time, and vacuum degree by means of the usual method without using the nozzle. The thickness of metallic film was about 5~6A. In this case the fine extreme point of tungsten wire of 0.2-mm diam was burned out, as the evaporation basket could not be used. Shadowing angle was $\tan^{-1}\frac{1}{5}$. Figure 1 is a positive image of carbon black by shadow-casting. In this figure (a), (b), and (c) are electron micrographs in the cases of vacuum degree of 0.08 mm Hg, 0.02 mm Hg, and 10^{-4} mm Hg, respectively, and (d) is that in the case of a vacuum of 2×10^{-5} mm Hg and of very slow evaporation after preheating for 10 hours. From this, it is seen that longer preheating time and better vacuum will result in a tendency for smaller aggregation, and shadow-casting becomes possible.

A similar experiment was made with Cr, which has been in general use. In this experiment the shutter was used, and evaporation of Cr was made at the evaporating temperature after the preheating at the higher temperature. The thickness was about 20A. Figure 2 is a positive image of Cr-shadowing. In this figure (a) is the electron micrograph in the case where Cr was evaporated at the evaporation temperature after preheating for 30 minutes and the vacuum degree was 5×10^{-6} mm Hg and 2×10^{-4} mm Hg before and after evaporation, respectively. On the other hand, (b) is the case where Cr was evaporated at the evaporating temperature after preheating for 2 hours. Shadowing angle of (a) and (b) are $\tan^{-1}\frac{1}{5}$ and $\tan^{-1}\frac{1}{8}$, respectively. It is known from this figure that the better the vacuum is, the better the shadowing will be. Also it is suggested that the method of using the shutter is very effective for shadowing.

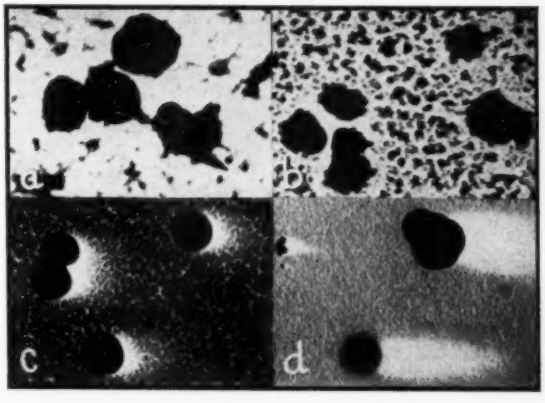


FIG. 1. Positive images of carbon black by W-shadowing in the cases of various vacuum degrees. (20,000×)

[1] R. C. Williams and R. W. G. Wyckoff, J. Appl. Phys. **15**, 712 (1944); **17**, 23 (1946).
[2] R. C. Williams and R. C. Backus, J. Appl. Phys. **20**, 98 (1949).

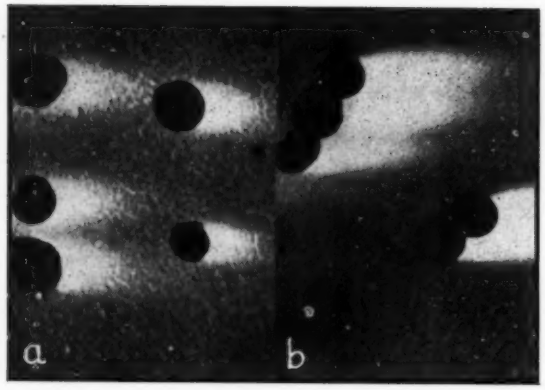


FIG. 2. Positive images of carbon black by Cr-shadowing in the cases of various vacuum degrees. (20,000×)

III. METALLIC SHADOW-CASTING APPARATUS WITH A NOZZLE SYSTEM AND THE METHOD OF USING IT

The metallic shadow-casting hitherto in use consists of coating the specimen obliquely with a very thin film of metal prior to the electron-microscopic examination. The coating is accomplished in an auxiliary vacuum chamber by means of a filament containing the metal to be deposited. As an application of shadowing, the methods of previous rotational shadowing[3] for obtaining the continuous structure of a substratum prior to the usual shadowing has been used. The double shadowing[4] for increasing the quantitative accuracy of shadowing and the portrait shadowing method,[5] wherein a metallic thin film is previously deposited in the region of shadow for increasing the image contrast and for preventing the charging up of the part of shadow due to intense electron beam, has been proposed. Any modification of the original shadowing apparatus, however, has not been made since Williams and Wyckoff.[1] In reality, ideal evaporation does not occur in vacuums of $10^{-4}\sim10^{-5}$ mm Hg, and shadowing materials are limited to a few metals. The shadowing technique is not simple, and shadowing quality of specimen is not always perfect. To minimize the aforementioned defects, metallic shadow-casting apparatus with a nozzle system can be used as follows:

Figure 3 shows the apparatus. In this figure *A* is an evaporation basket of 0.5-mm diam tungsten wire and *B* is a specimen supporter. *I* is a bell jar and *J* connects to the vacuum pump. These parts constitute a standard shadowing apparatus. The present device, however, has a shutter *C* of iron, a nozzle *D* of glass, and a cover plate *E* between *A* and *B*. The shutter *C* can be opened by switching off the electric current through the electromagnet *K*. By using rails *G* and *H*, the cover plate can be moved to change distances from the evaporation basket and heights from the base, respectively. The section of the nozzle is large at the entrance and narrow at the exit as shown in Fig. 3. The narrower and the longer the nozzle, the better the shadowing will be, but the size and length of the nozzle should be changed according to the hardness or easiness of evaporation. In this experiment the dimensions of the nozzle were chosen from the maximum 3-mm diam by 1.5-mm diam and 22 mm, to the minimum 1.5-mm diam by 0.7-mm diam and 38 mm as diameter at the entrance, diameter at the exit, and the length, respectively. The specimen supporter *B* is covered with a glass vessel *F* for the purpose of eliminating metallic vapor other than that from the nozzle. For a comparison of the case of not using a nozzle with the case of using a nozzle, a hole of 26 mm×16 mm was made at another place on a cover plate, and a shutter was also placed before this hole. A specimen was placed farther away than in the case of using a nozzle and was also covered with a glass vessel. These parts of the apparatus are necessary only in the case of a comparison experiment and so are unnecessary in the case of usual shadowing.

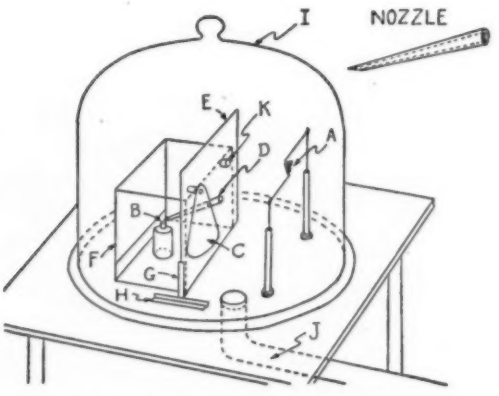


FIG. 3. Shadowing apparatus using a nozzle.

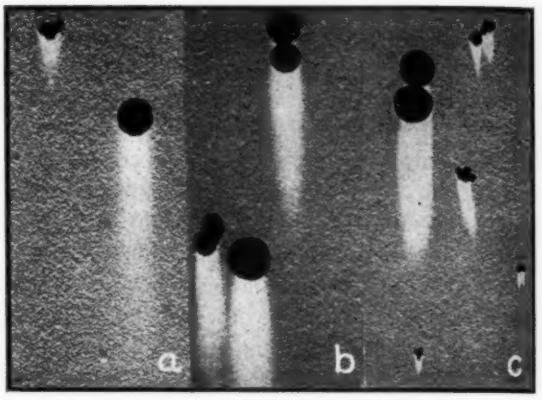


FIG. 4. Positive images of carbon black by Cr-shadowing by the methods hitherto used (a) of using the shutter, (b) and (c), of both using the shutter and the nozzle. (20,000×)

The method of using this apparatus is as follows: First, the shadowing metal (about three times the quantity for the usual shadowing method) is put in the tungsten basket which has previously been well out-

[3] F. Heinmets, J. Appl. Phys. **20**, 385 (1949).
[4] R. C. Backus and R. C. Williams, J. Appl. Phys. **20**, 224 (1949).
[5] Delbert E. Philpott, J. Appl. Phys. **22**, 982 (1951).

gassed by preheating in high vacuum. After the vacuum in the bell jar reached about 10^{-5} mm Hg, the evaporation metal is heated up above its evaporating temperature and is well outgassed with the shutter closed. Though the metal has already evaporated by this time, metallic vapor will not reach the specimen because of the closed shutter. When a pressure better than 10^{-4} mm Hg is maintained following this outgassing, the current to the filament is switched off and the shutter is opened. Then the metal is evaporated at a lower temperature than the outgassing one. In this case control of the thickness of deposited metallic thin film comes into question, for there are shadowing metals for which the thickness of metallic film can easily be determined and others for which it cannot be determined from the coloring of a glass placed under the specimen holder. By repeating the shadowing a few times, however, an exactly suitable thickness of the deposited metallic thin film can easily be found, for the shadow of the specimen never fails to be cast by means of this method even if the thickness of the deposited film is too thin or too thick. This is one advantage of the apparatus with a nozzle system.

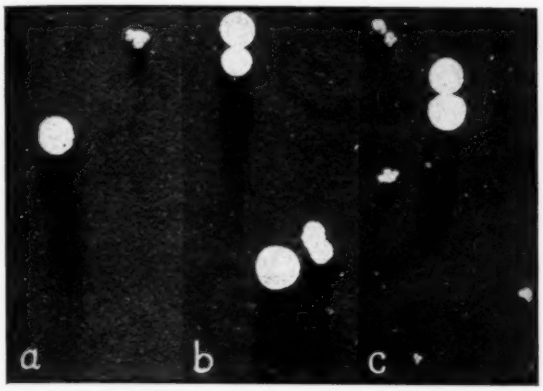


FIG. 5. Negative images of carbon black by Cr-shadowing by the methods hitherto used (a) of using the shutter, (b) and (c), of both using the shutter and the nozzle. (20,000×)

An "Hitachi" magnetic electron microscope (made in Japan) was used and electron micrographs were taken at 4500 times electron-optically and optically enlarged to 20,000 times, excepting a few cases.

IV. EFFECT OF THE SHADOWING APPARATUS WITH A NOZZLE SYSTEM

In order to confirm the effects of shutter and of nozzle separately, three methods hitherto used (using the shutter and using both the shutter and the nozzle) were examined comparatively using Cr-shadowing of carbon black on collodion foils. These results are shown in Figs. 4 (a)–(c) and Figs. 5 (a)–(c), respectively. Figures 4 (a)–(c) are positive images and Figs. 5 (a)–(c) negative images. From these, it is seen that shadowing quality is considerably improved by outgassing only and the nozzle system is very effective for the metallic shadowing. Figure 6 is a Cr-shadowed negative image of "Micronex" (very fine carbon black) and shows very delicate shadows. Even without using a nozzle a shadowing metal such as Cr can cast the shadow, but a metal such as Sb cannot be used as a shadowing material. Figure 7 (a) is a positive image without using a nozzle for Sb, and (b), (c), and (d) are also positive images of Sb-shadowing in the case of using a nozzle. In the case of (c) the nozzle used is narrower and longer than in the case of (b). From these, the effect of the nozzle can be seen distinctly. Figure 7 (d) is an electron micrograph for the case of outgassing of the collodion foil by the previous electron bombardment for the purpose of determining the effect of the substratum. In (d) deposited particles are fine and the shadowing quality is better, but many aggregation specks appear over all the visual field. An extreme case is seen in an example of Zn. Except in the case of the well-outgassed collodion foil by the previous electron bombardment, a deposited film cannot be obtained at all in this case. Figures 8 (a) and (b) are positive images in the case of a well-outgassed substratum using Zn with and without the nozzle, respectively. Shadows are obtained, but aggregated large particles appear at many places owing to the easiness of aggregation.

FIG. 6. Negative image of Micronex by Cr-shadowing. (45,000×)

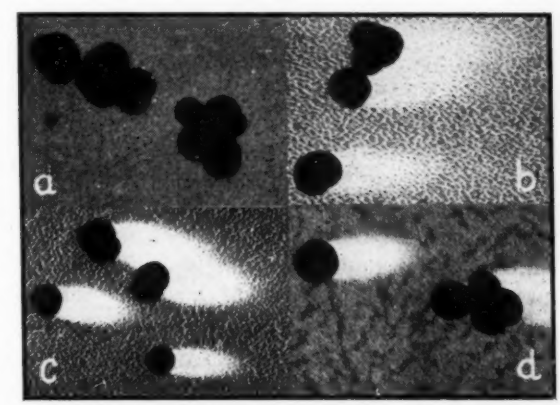


FIG. 7. Positive images of carbon black by Sb-shadowing without and with using the nozzle. (20,000×)

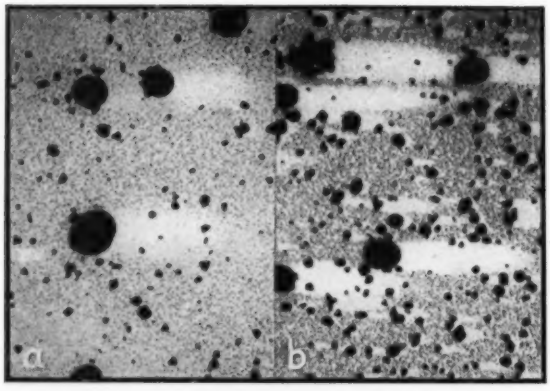


FIG. 8. Positive images of carbon black by Zn-shadowing in the case of well-outgassing of collodion foil without and with using the nozzle. (20,000×)

As mentioned above, if the nozzle is used, the shadowing quality of metals hitherto used as shadowing materials becomes better and the metals which could not be used as shadowing material may be used due to grain refinement. In addition, the nozzle system has the following benefits: (1) Without using a vacuum gauge it is found that in the case of poor vacuum the evaporated metal passes through the nozzle with difficulty; (2) since a point source of evaporation is no longer necessary, the operation is very simple and evaporated metals can be used in the form of either lump or powder.

V. METALLIC SHADOWING ELEMENTS

It was already reported by Williams and Backus[2] that U was the best shadowing metal, that Au was easily evaporated but it had a defect in that it was easily aggregated, that Cr was comparatively a suitable shadowing element, and that Mn could also be used. Excepting a few examples, Cr- or Au-shadowing was used generally. In the case of using the nozzle system, however, metals over a considerably wide range besides the aforementioned metals can be used as the shadowing metals and the mutual comparison of shadowing quality of evaporated metals becomes easier. The shadowing quality was compared here for many different metals and the order of shadowing quality determined. In this case, the thickness of the deposited thin metallic film and adsorbed quantity of gases on the collodion foil come into question. The thickness of the thin film which gives a suitable contrast on the electron micrograph was chosen as the standard thickness for all metals and the collodion foil as substratum was used in the state as it was made. Accordingly, it will be risky to discuss quantitatively results obtained, but the conditions will allow a qualitative comparison of the shadowing qualities of metals. If one aims at a more quantitative measurement of shadowing qualities apart from the aforementioned two conditions, the method of using Dow Latex polystyrene particles, which was proposed by Backus and Williams[4] and applied to the study of polished metallic surface by I-Ming Feng,[6] should be adopted, though it will not be tried here.

Positive images and negative images of carbon black shadowed by various metals without and with the nozzle are shown in Figs. 9* and 10, respectively. Also from the standpoint of practical use the results obtained are tabulated in Table I.† From these figures it is found that the shadowing quality becomes better in the case of using the nozzle. From the third and fourth columns of Table I classifying the elements of periodic chart into three groups‡ of "*a*" and VIII (A), "*b*" (B), and elements not belonging to "*a*" or "*b*" (C), it is found that in all groups the larger the heat of sublimation is, the better the shadowing quality is. It should be noticed that this rule differs from the result obtained by Levinstein[7] on the aggregation qualities of deposited films.

As most metals listed in this table have good shadowing quality in the case of using the nozzle, a suitable metal may be chosen according to the practical necessary condition. In these metals W and Mo are superior shadowing metals for casting a fine shadow and for quantitatively determining the size or form of specimen, but their evaporation technique is very difficult. Hence Th or Zr is to be chosen as the better shadowing metal because of the comparatively slight thickness of the metallic film. In particular, Zr is a very desirable shadowing metal because of the good vacuum at the evaporating time due to its pronounced gettering action.

The dependence of shadowing quality upon evaporation rate was investigated. The examples of Mn belonging to "*a*" and Sn belonging to "*b*" are shown in Fig. 11, A and B, respectively. In this figure (a) is a negative image in the case of smaller evaporation rate at lower temperature and (b) is a negative one in the case of larger evaporation rate at higher temperature. It was known that "*a*" and group VIII showed a better shadowing quality with higher temperature and larger evaporation rate, but "*b*" was better with lower temperature and smaller evaporation rate.

---

[6] I-Ming Feng, J. Appl. Phys. **22**, 820 (1951).

* In Fig. 9 positive images of many elements in the case of not using the nozzle are distinguished between two cases, namely, with and without using the shutter. Zr, Ti, Fe, Cr, Mn, and Ag belong to the latter case and the other elements belong to the former case.

† In this table the values of heat of sublimation were taken from the table by F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Chapter 1, on all elements excepting Si. The value of the heat of sublimation at 2355°C is quoted as the one of Si.

‡ This classification refers to the periodic chart of the *International Critical Tables of Numerical Data, Physics, Chemistry, and Technology*, Vol. I, published for the National Research Council (McGraw-Hill Book Company, Inc., New York, 1926) and of the *Landolt-Börnstein Physikalisch-Chemische Tabellen*, Von Walter, Roth, and Scheel (Spamerschen Buchdruckerei, Leipzig, 1923).

[7] H. Levinstein, J. Appl. Phys. **20**, 306 (1949).

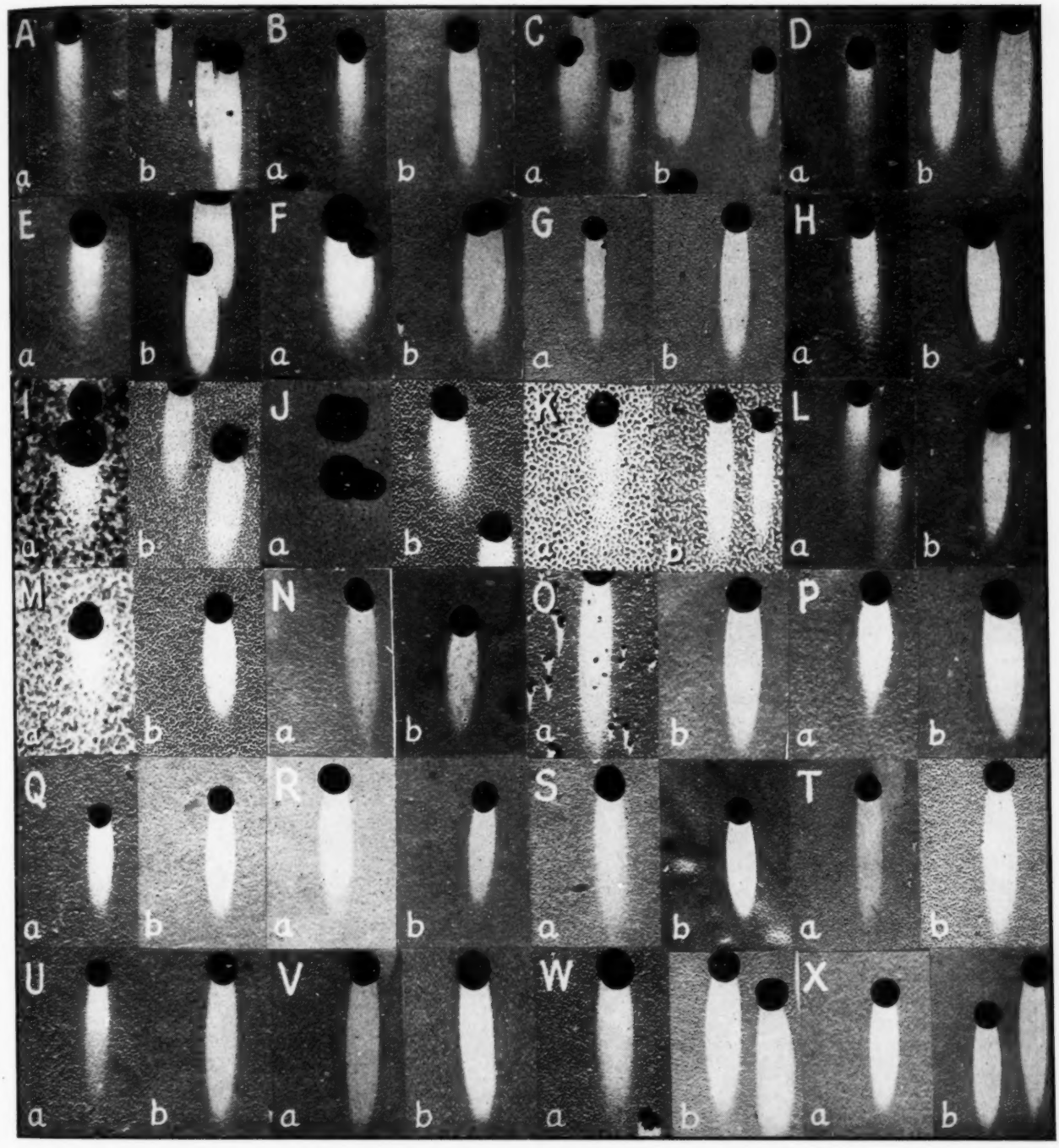


FIG. 9. Positive images of carbon black by various metallic shadowings without (a) and with (b) using the nozzle. *A*: Cr, *B*: Mo, *C*: W, *D*: Mn, *E*: V, *F*: Ni, *G*: Co, *H*: Fe, *I*: Bi, *J*: Sb, *K*: Pb, *L*: Cu, *M*: Ag, *N*: Ti, *O*: Zr, *P*: Th, *Q*: Yt, *R*: Pt, *S*: Au, *T*: Sn, *U*: Ge, *V*: Be, *W*: Al, and *X*: Si. (20,000×)

VI. CONSIDERATION OF RESULTS OBTAINED

In the case of using the nozzle system with shutter, a uniformly deposited film and a fine shadow can be obtained. It may be attributed to the following causes: (1) As shadowing elements may be well outgassed if the shutter is used, deterioration of the vacuum is prevented during the period of evaporation and fine particles are produced on specimen. (2) Since a perfect vacuum cannot be obtained even though the shutter is used, evaporated particles having different directions and velocities impinge on the specimen. The purpose of the nozzle is to minimize these effects, so if the nozzle is used, a well-defined shadow and uniformly deposited metallic film having only very few centers of granular aggregation can be obtained. (3) It may be considered that owing to reflections from the inner wall of the

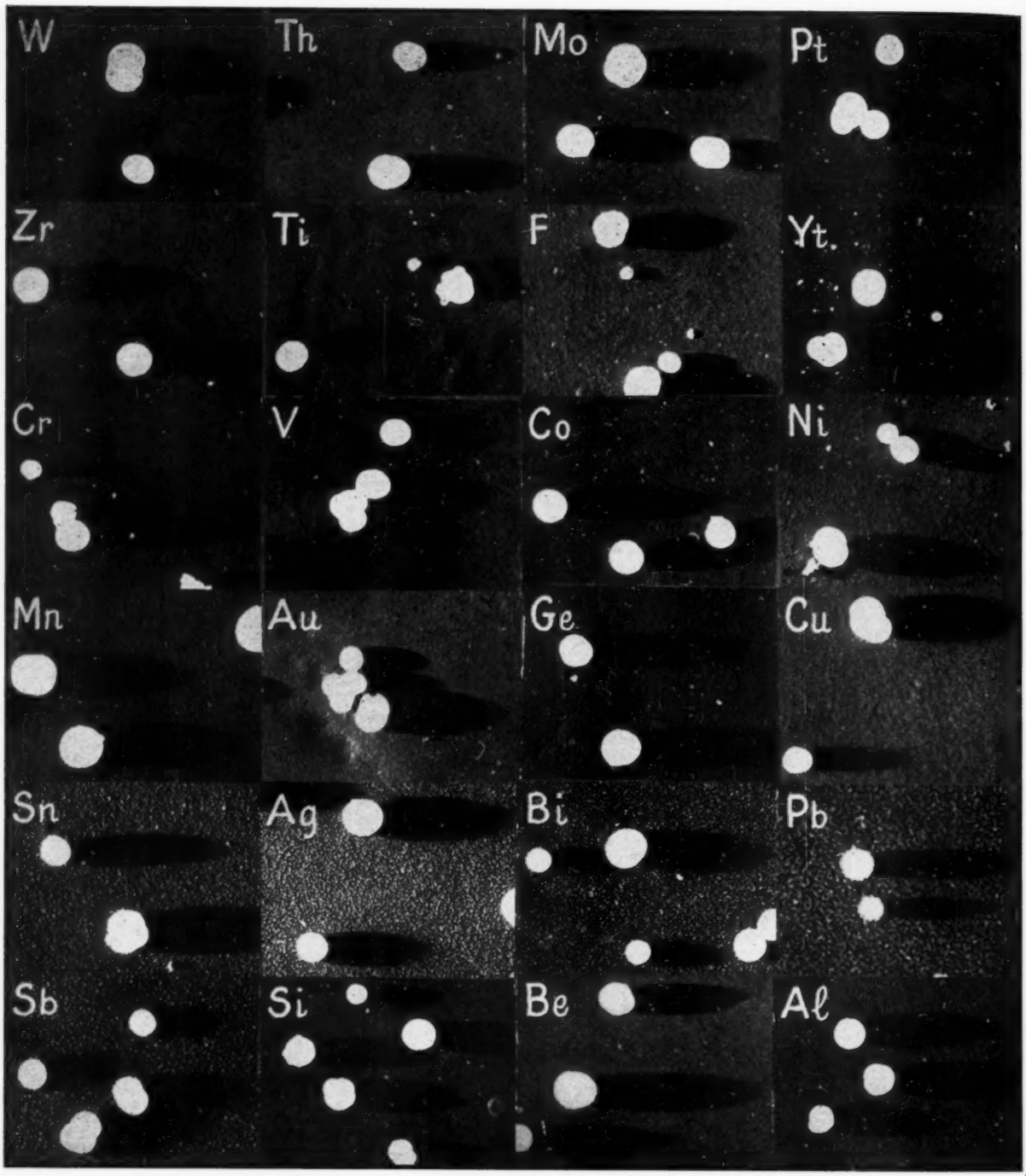


FIG. 10. Negative images of carbon black by various metallic shadowings. (20,000×)

nozzle, particles coming through the nozzle have very slow velocities and so do not migrate large distances on the condensing surface. (4) By using the cover plate, secondary effects of evaporation can be eliminated and heat radiation during evaporation is incident on the foil supporting the specimen only through the nozzle, so that the substratum experiences only a slight temperature rise. (5) The effect of the previous electron-bombardment on the substratum may be attributed to removing the adsorbed gases. Owing to this effect the adhesion force between deposited elements and substratum increases and aggregation becomes difficult.

A new rule for the order of aggregation of deposited metallic thin film was proposed experimentally. In the three groups already mentioned the larger the heat of sublimation is, the better the shadowing quality. A complete physical interpretation, however, must wait. If this rule is correct, it may be inferred that rare-earth

TABLE I. Shadowing elements and their shadowing qualities.

Group	Metal	Rank	Heat of sublimation (kcal/mole at room temp)	Distance (cm)	Melting point (°C)	Hardness or easiness of evaporation	Before or after melting	Alloying with W	Hardness or easiness of discrimination	Hardness or easiness of removal	Remark
A	W	1	210	4	3400	hard	before		hard	−	
	Th	2	177	5	1842	slight hard	before	none	hard	−	powder
	Mo	3	160	4.5	2622	hard slight	before	small	hard	−	powder
	Pt	4	127	5	1773	hard slight	same time	large	hard slight	±	lump
	Zr	5	110	5.5	1857	hard slight	same time	mediate	hard	−	powder
	Ti	6	100	5	1800	hard slight	same time	small	hard	−	powder
	Fe	7	94	5	1530	hard slight	after	large	easy	±	powder
	Yt	8	90	5	...	hard	same time	none	hard	−	powder
	Cr	9	88	5	1805	easy	before	none	easy slight	−	lump
	V	10	85	5	1800	easy slight	before	none	hard	−	lump
	Co	11	85	5	1489	hard slight	after	large	easy	±	lump
	Ni	12	85	4.5	1455	hard	after	large	easy	±	lump
	Mn	13	74	5	1247	easy	before	none	easy	−	lump
B	Au	1	92	5	1063	easy	after	none	easy	±	lump
	Ge	2	85	5	958.5	easy	after	none	easy	−	lump
	Cu	3	81.2	5	1080	easy	after	none	easy	±	lump
	Sn	4	78	5	231.8	easy	after	none	easy	+	lump
	Ag	5	68	5	960	easy	after	none	easy	+	lump
	Bi	6	47.8	4.6	271	easy	after	none	easy	+	lump
	Pb	7	47.5	4.4	327	easy	after	none	easy	+	lump
	Sb	8	40	3.5	680	easy	after	none	easy	+	lump
C	Si	1	(94.3)	5	1420	easy	same time	mediate	easy	−	lump
	Be	2	75	5	1278	easy	same time	none	easy slight	−	lump
	Al	3	55	5	659	easy	after	small	hard slight	−	lump
	Li	4	39	5	179	easy	same time	small	hard	−	lump

metals have not such good shadowing quality, and U is the best shadowing metal as found by Williams, though it was not tried in this experiment.

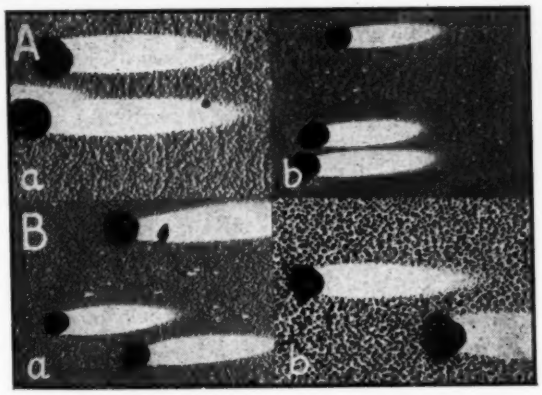


FIG. 11. Positive images of carbon black by Mn-shadowing (A) and Sn-shadowing (B) in the cases of smaller evaporation rate (a) and of larger evaporation rate (b). (20,000×)

The transition metals Mn, V, Cr, Mo and W, and Th belonging to "*a*" of periodic chart were evaporated at temperatures below their melting points as seen in Table I, and a well-defined shadow was obtained. On the other hand, Au, Cu, Ag, and Sb belonging to "*b*" of the periodic chart were evaporated at temperatures above their melting points, and aggregation seems to occur readily. However, it is very interesting to note that the iron group has comparatively good shadowing quality notwithstanding its evaporation at temperatures above the melting point.

VII. ACKNOWLEDGMENT

The author wishes to express his appreciation to Professor R. C. Williams and Dr. N. Higashi for their kind interest in this investigation. He would also like to express his thanks to Mr. K. Yada, Mr. S. Takahashi, and Mr. S. Takai for their helpful assistance.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Reflection of Microwaves from Metal-Plate Media*

J. J. Brady, M. D. Pearson, and S. R. Peoples
Department of Physics, Oregon State College, Corvallis, Oregon
(Received March 10, 1952)

The theory of Carlson and Heins and the extended theory of Lengyel has been verified experimentally for oblique angles of incidence. Both single surface and slab measurements were performed in the 3-cm band. Single surface measurements were made by using tapered wooden wedges as absorbing materials to eliminate back surface reflections. The effect of plate thickness was studied for five different thicknesses. The phase and amplitude of the reflection coefficient of a metal-plate slab were obtained for three angles of incidence, 10°, 25°, and 35°, and for values of $2a/\lambda$ ranging from 1.10 to 1.30.

INTRODUCTION

ELECTROMAGNETIC waves confined between conducting plates which are parallel to the electric vector and spaced more than a half-wavelength apart have a phase velocity greater than their free-space velocity. A row of such parallel plates constitutes a refractive medium with an index of refraction less than unity. A medium of this kind, with the proper profile, can be used to focus waves much in the same manner as a lens of dielectric material.

The purpose of this study is to verify experimentally the theory of Carlson and Heins[1] as well as the extended work of Lengyel[2] regarding the reflection coefficient of metal-plate media. Lengyel has found an expression for the reflection coefficient when a single diffraction beam is present and has compiled tables and graphs giving coefficients which describe electromagnetic behavior at the surface of metal-plate media. The Carlson-Heins theory is applicable only in the presence of a single reflected beam. Hence, the major part of this present work is concerned with Lengyel's results and his notation will be used throughout. The surface of the medium is normal to the plates and the E-vector is parallel to their edges. This is a special case of the general theory. Ruze and Young[3] made measurements in a closed system while the experiments described here were made in "free space."

SINGLE SURFACE MEASUREMENTS

Slabs 1 and 2 were both constructed of square wave guide sections of equal lengths ($8\frac{3}{4}$ inches) containing wooden matching wedges. An individual guide section and matching wedge are shown in Fig. 1. The sections were assembled and held together by a wooden frame. The assembly for Slab 1 is shown in Fig. 2. The depth of the slab, d, the plate separation, a, the plate thickness, t, and the area of the slab surface, A, are given in Table I.

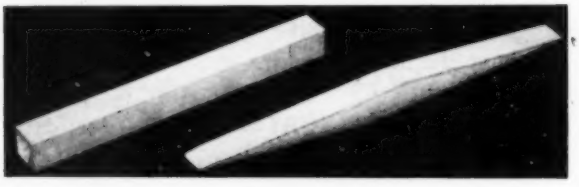


FIG. 1. Guide section and matching wedge.

Experiments with Slab 1

A Sperry 2K39 klystron was the source of radiation. It operated in the wavelength band from 3.2 to 3.7 cm, providing a $2a/\lambda$ ratio from 1.15 to 1.35. A bolometer (Sperry Barretter) served as a detector in conjunction with a Radiation Laboratory Mark IV amplifier and a Ballantine voltmeter. Provision was made for recording the data on the tape of an automatic recorder. The vertical deflection of the trace on the tape was proportional to power received.

A spectrometer-like apparatus was constructed with a fixed transmitting paraboloid (8 inches diameter with dipole feed) and a rotating arm carrying the receiver horn. This apparatus is shown in Figs. 3 and 4.

The movable arm was rotated slowly by a motor through a gear reduction drive and the received power was recorded on a tape moving in synchronism with the receiving horn. Precautions were taken to insure accurate identification of the angular position of the horn from the markings on the recording tape. The horn could swing over an arc of 340° leaving 10° clearance on each side of the transmitting antenna.

In every experiment the slab was set for a fixed angle of incidence, and the receiver arm was rotated to obtain a pattern of the reflected and transmitted radiation. When a reflection pattern was taken, the slab was so positioned that the axis of rotation was in its front surface. For transmission measurements the slab was positioned with the axis in its black surface. In order to obtain both transmission and reflection measurements on the same tape, the instrument was stopped at the proper point, and the slab was moved forward or back, keeping its orientation unchanged. This displacement of the slab was always accompanied by a corresponding displacement of the transmitter necessary to preserve the transmitter to slab distance.

Figure 5 shows samples of the record tape obtained

* The early part of this work was supported by NRL and the latter part by the ONR. It was published with the approval of the Monographs Publication Committee, Oregon State College, Research Paper No. 203, Department of Physics, School of Science.

[1] F. Carlson and A. E. Heins, Quart. Appl. Math. 4, 313 (1947).
[2] B. A. Lengyel, J. Appl. Phys. 22, 265 (1951).
[3] J. Ruze and M. Young, J. Appl. Phys. 22, 277 (1951).

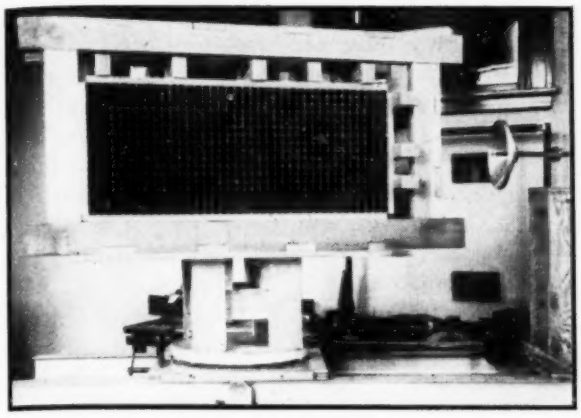


FIG. 2. Wave-guide assembly (Slab 1).

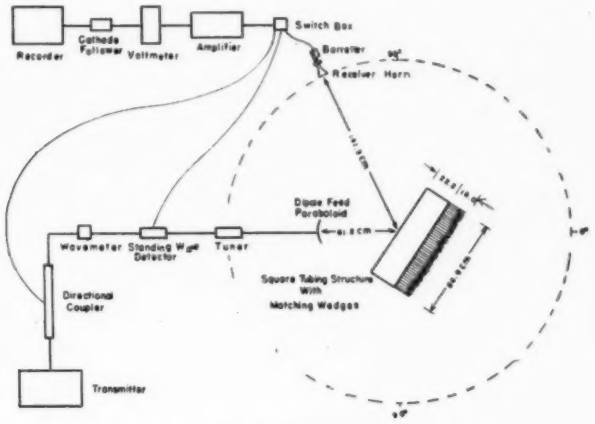


FIG. 3. Diagram of apparatus with Slab 1.

from a parallel plate slab of the same dimensions as Slab 1, but having no matching wedges. Hence, transmission data are recorded. The transmitting antenna was always at 180° on the scale shown. The letters F and B refer to the front and back surfaces, respectively, indicating which one contained the axis of rotation. The position of the angular zero is marked by the sharp dip of the trace, caused by the action of a microswitch, which connected the recorder to a fixed reference of 0.3 volt at the instant the spectrometer arm passed the null position.

It is in the nature of these experiments that rather wide lobes are obtained on the record tape, extending 10 to 15 degrees on each side of their maxima. These maxima are poorly defined because of the fine structure of the lobes. For this reason the experimental definition of reflection and transmission coefficients was not based on the comparison of maxima, but on the areas under the curves. The amplitude squared (intensity) of the reflection coefficient was defined experimentally as the ratio of the area under a reflected lobe to the area under the main lobe received with the slab removed from the path of the radiation. This definition required a separate determination of the latter area for each wavelength used. In addition, all data were normalized to a constant monitor reading. As a check on the normalization process a solid metal plate was placed in front of the slab, and its reflected pattern was obtained. When the area under the reflected lobe of this pattern was used as the denominator of the expression for the reflection coefficient, the values agreed within a few percent with those obtained by the previous method.

When the angle of incidence θ is less than the limiting angle θ_L defined by the equation, $1+\sin\theta_L=\lambda/a'$, the theory predicts the presence of a single reflected beam at an angle equal to the angle of incidence. More precisely, the incident and the reflected wave normals are oriented at θ and $\pi-\theta$, respectively. This is the case of specular reflection, so well known in geometrical optics. When $\theta\geqq\theta_L$, then, in addition to the wave propagating in the direction $\pi-\theta$, diffracted waves will be present. In the cases of practical interest there is only one diffracted beam; the direction θ' of its wave normal is specified by the equation $\sin\theta+\sin\theta'=\lambda/a'$, where $a'=a+t$.

The theory is based on the assumption that the plates have zero thickness, therefore plate separation a and plate spacing a' are identical. In experimental work these quantities are slightly different. While propagation between the plates depends on a, diffraction grating effects depend on a'.

It has been shown theoretically that the intensity of specular reflection decreases gradually as θ varies from 0 to θ_L, and then rises very sharply. The angle of incidence for which the reflection coefficient is a

TABLE I. Dimensions of metal plate slabs in centimeters.

Slab No.	A	d	a	t
1	84.5×35.7	22.17	2.143	0.079
2	17.8×17.8	22.17	2.143	0.079
2*a*	18.7×22.2	22.17	2.143	0.206
2*b*	19.6×22.2	22.17	2.143	0.333
2*c*	21.5×22.2	22.17	2.143	0.613
2*d*	24.5×22.2	22.17	2.143	1.045
3	20.0×20.0	10.00	1.920	0.064

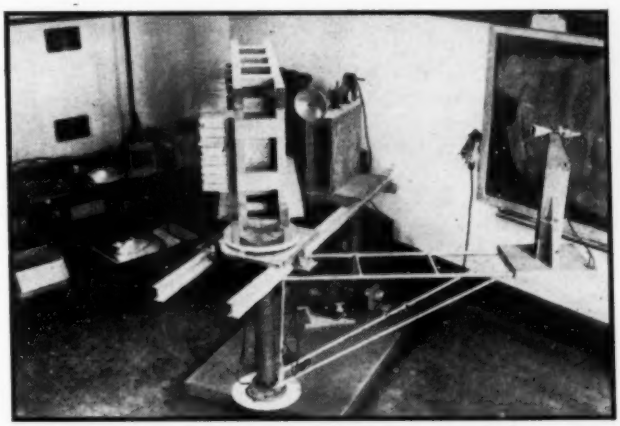


FIG. 4. Spectrometer with Slab 1.

minimum should then coincide with the angle for which the diffracted beam first appears. The observed angular position of the specular reflection coincides with the position predicted by the theory. The diffracted beam, however, makes its first appearance for an angle of incidence smaller than θ_L. This fact is brought out in Fig. 6 where the minima of the intensity of specular reflection are also shown. The apparent discrepancy is understandable in view of the fact that the experiment is not performed with parallel beams; therefore the angle of incidence varies over the surface of the slab. The diffracted beam will appear when conditions for its appearance are satisfied on any one part of the reflecting surface. Moreover, when the diffracted beam first appears it has a broader power pattern than the incident beam, because of the small apparent size of the reflector when viewed from the direction of the diffracted beam. This fact further enhances the early appearance of the diffracted signal in the receiver. Finally, from the asymmetrical nature of the curves representing the reflection coefficient as a function of the angle of incidence, it follows that the minimum of the reflected intensity will be shifted toward smaller mean angles of incidence when a divergent incident beam is used in place of a parallel beam.

The reflection coefficient of a single surface is defined by Lengyel as $r=\rho e^{-i\rho'}$. He has shown that

$$\rho=\frac{\cos\theta-n\cos\theta+\cos\theta'}{\cos\theta+n\cos\theta-\cos\theta'},$$

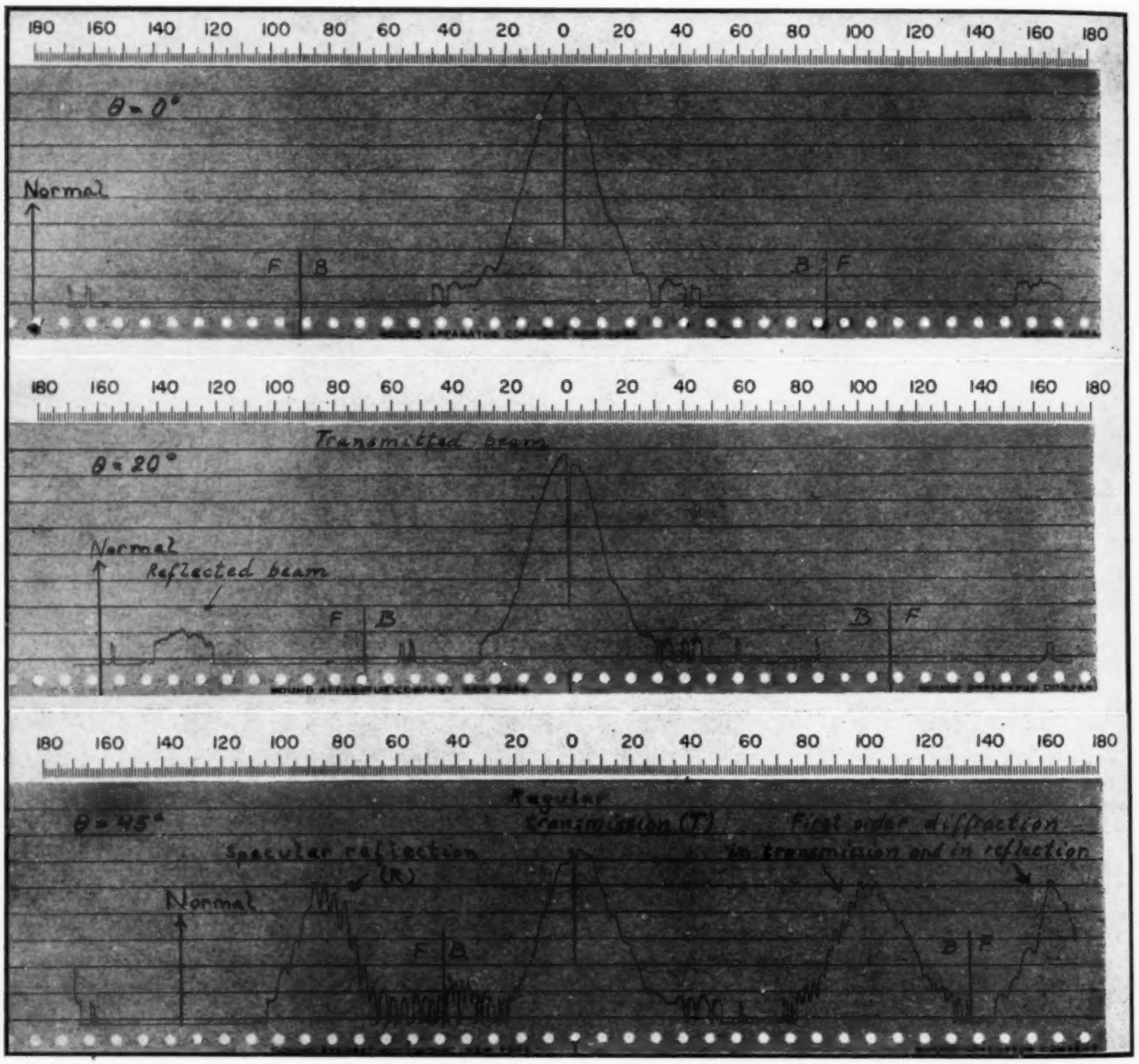

180 160 140 120 100 80 60 40 20 0 20 40 60 80 100 120 140 160 180
θ = 0°
Normal
F B
B F
180 160 140 120 100 80 60 40 20 0 20 40 60 80 100 120 140 160 180
Transmitted beam
θ = 20°
Normal
Reflected beam
F B
B F
180 160 140 120 100 80 60 40 20 0 20 40 60 80 100 120 140 160 180
θ = 45°
Regular transmission (T)
First order diffraction in transmission and in reflection
Specular reflection (R)
Normal
F B
B F

FIG. 5. Samples of records from automatic recorder.

where θ is the angle of incidence, n, the index of refraction, is given by $n=[1-(\lambda/2a)^2]^{\frac{1}{2}}$, and $\theta'=\sin^{-1}(\lambda/a'-\sin\theta)$ is the angle the diffracted beam makes with the surface normal. Representative values of ρ measured for Slab 1 are shown in Fig. 7 where the two solid curves are computed from the above theoretical formula, using both a and a' in the expression for θ'.

For angles of incidence of 10 to 25 degrees, the measured value of the reflection coefficient is consistently below the value predicted by the theory (solid curves). This apparent discrepancy is, no doubt, due to the fact that a parallel beam was not used in the experiment. It is likely that the measurements are subject to a systematic error which is more apparent in the case when ρ is small than when ρ is large.

Lengyel[4] has shown that when the angle of incidence is equal to arc cos$n$, the first-order diffracted beam is reflected back into the transmitter and an interesting equidistribution of power takes place, one-half of the energy being transmitted into the metal-plate medium, the other half being divided equally between the two reflected beams. As L. J. Chu† of M.I.T. pointed out, in this case the transmitted wave consists of two plane waves inclined at $\pm\theta$ to the z-axis. Within the medium the electric field is $i\exp(iknz)\sin\pi x/a$. This may be written as the sum of two plane waves,

$$\tfrac{1}{2}\exp ik(z\cos\theta+x\sin\theta)-\exp ik(z\cos\theta-x\sin\theta),$$

which satisfy the boundary conditions on the metal plates. They will emerge at the back surface of the metal-plate medium without reflection as two plane waves of equal amplitude each carrying one-fourth of the flow of energy incident on the front surface.

The distribution of power in the presence of diffracted beams is exemplified in Fig. 8. The points represent measurements using the parallel plate model of Slab 1 without wedges of the ordinary reflected and transmitted beams together with the diffracted transmitted beam. The measurement of the fourth beam had to be omitted because, for most angles of incidence shown, it was directed near the transmitter. Chu's check point where $\theta=\cos^{-1}n$ and all beams have relative power values of 0.25 is indicated on the two diagrams with an X. Similar diagrams obtained for other values of the index of refraction provide a verification of Chu's prediction.

The amplitude ρ of the reflection coefficient has been obtained for angles of incidence between 10° and 60° over the range $2a/\lambda=1.15$ to 1.35. These agree essentially with the theory, although, because of experimental difficulties, a quantitative agreement was not obtained in all cases. In particular, it has been shown that the intensity of specular reflection decreases slowly with the angle of incidence until a critical limiting angle is reached. Beyond this angle the conditions for the formation of a diffracted beam are satisfied, and the intensity of the specular reflection rises sharply.

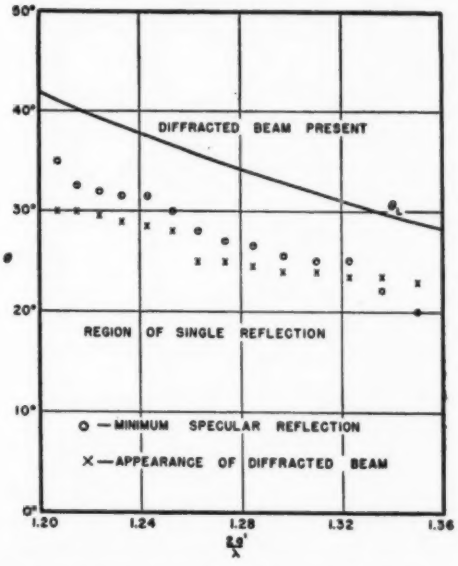


FIG. 6. The least angle of incidence for diffraction.

THE EFFECT OF PLATE THICKNESS

Experiments with Slab 2

The pulsed magnetron of an APS-4 radar served as the source of radiation. It operated at the fixed wavelength $\lambda=3.2$ cm. Detection of the signal was accomplished as before except that the automatic recorder was omitted and peaks of lobes were read from the Ballantine voltmeter. The spectrometer was used with the fixed transmitting paraboloid (5-foot diameter with horn feed) about 70 feet from the center of rotation and the receiving horn about 10 feet away from the metal-plate medium. The dimensions of Slab 2 are given in Table I.

Since the maxima were more clearly defined, it was now satisfactory to compare intensities of the peaks of lobes. Normalization was accomplished by means of a metal plate having its area identical to the area of the face of the slab. The experimental value of ρ was then determined from the square root of the ratio between the voltmeter reading for the slab and the normalizing metal plate.

As in the experiments with Slab 1, single surface measurements were made. To test the efficiency of the matching wedges, a check was made using the variable frequency klystron with the transmitter to center of rotation distance decreased and a smaller transmitting antenna substituted. The effectiveness of the wedges is clearly indicated by the contrast between the amplitude of the reflection coefficient *versus* $2a/\lambda$ curves with and without wedges shown in Fig. 9. The values for the theoretical curve of ρ *versus* $2a/\lambda$ at $\theta=25°$ were obtained from Lengyel's tables.[4] The other theoretical

[4] B. A. Lengyel, NRL Report 3534 (1949).

† At a conference held at the Naval Research Laboratory, Washington, D. C.

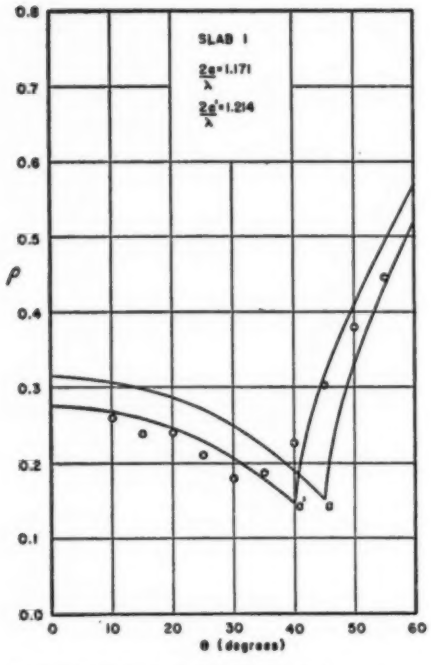


FIG. 7(a). Amplitude of reflection coefficient as a function of angle of incidence. (Solid line a': calculated from Lengyel's theory using center-to-center plate spacing; Solid line a: same theory using inside plate spacing.)

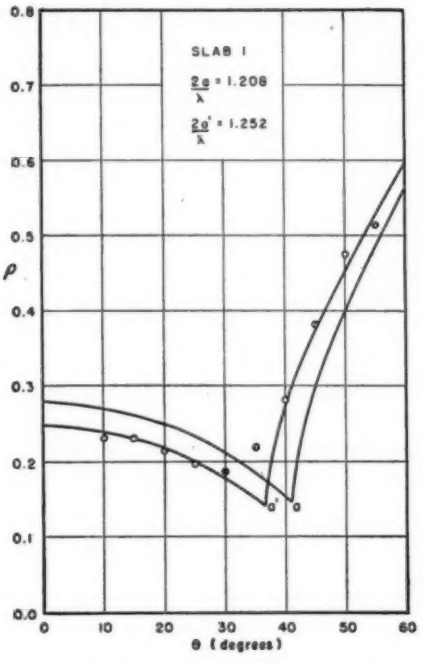


FIG. 7(b). Angle of reflection coefficient as a function of angle of incidence.

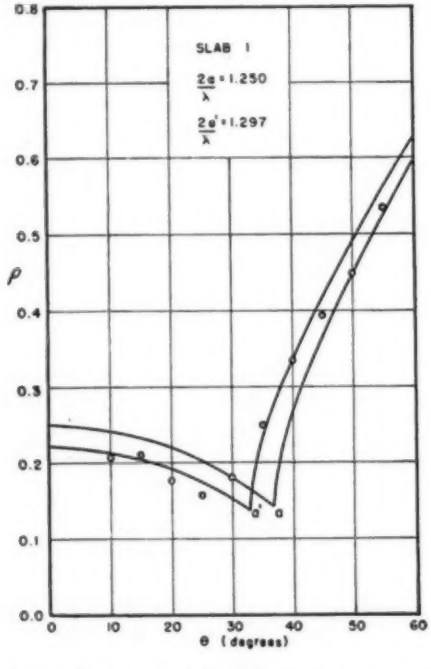


FIG. 7(c). Amplitude of reflection coefficient as a function of angle of incidence.

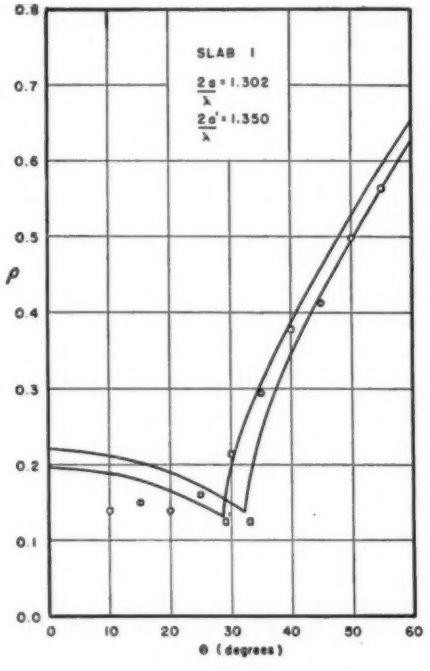


FIG. 7(d). Amplitude of reflection coefficient as a function of angle of incidence.

curve for $|R|$ *versus* $2a/\lambda$ was calculated from Lengyel's result for a slab

$$|R| = \frac{2\rho|\sin\psi|}{[(1-\rho^2)^2+4\rho^2\sin^2\psi]^{\frac{1}{2}}},$$

where $\psi=\rho''+2\pi nd/\lambda$. Here ρ'' is the phase of the back surface reflection coefficient of a slab of metal-plate medium and likewise was obtained from Lengyel's tables.[4]

Figure 10 shows ρ as a function of θ for $2a/\lambda=1.341$. Here as in Fig. 7, theoretical curves are plotted for both a and a'. For each value of θ, the average of 3 or

4 experimental determinations is plotted showing the limits of probable error. The plate thickness t was varied by inserting metal sheets between rows of wave guides. Figure 11 shows ρ *versus* θ at $2a/\lambda=1.341$, using slabs $2a$–d, whose respective values of t were increased in small increments and are given in Table I. A careful study of these curves reveals the fact that the experimental points follow the theory more closely, when a is used to compute the theoretical curve for $\theta<\theta_L$ and when a' is used for $\theta>\theta_L$. As has already been pointed out, propagation between the plates

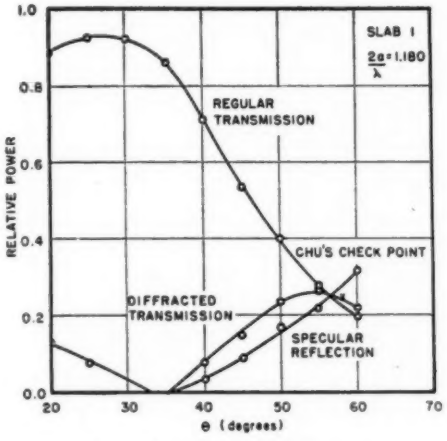


FIG. 8(a). Division of power.

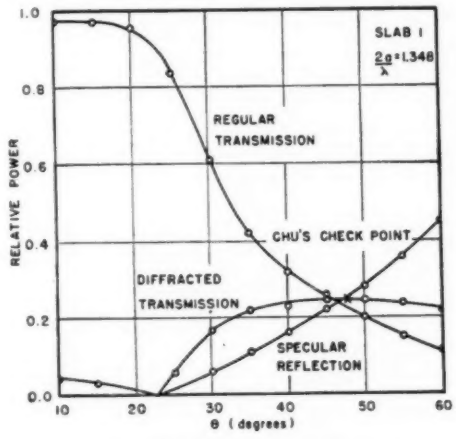


FIG. 8(b). Division of power.

depends on a, while diffraction grating effects depend on a'. In the case of slab $2d$ where $t=1.045$ cm and $a=2.143$ cm, $a'=3.188$ cm approaches the wavelength $\lambda=3.195$ cm. The ratio t/a is about $\frac{1}{3}$, which indicates that $\frac{1}{3}$ of the slab aperture is metal. In view of this fact it is rather surprising that experimentally $\rho=0.261$ at $\theta=10°$. Simple reasoning might lead one to expect the minimum value in this case to be $\rho=0.577$ corresponding to $\frac{1}{3}$ of the power being reflected. However, in this experiment the latter value of ρ was not reached until $\theta=40°$. No theory is yet available to account for these results.

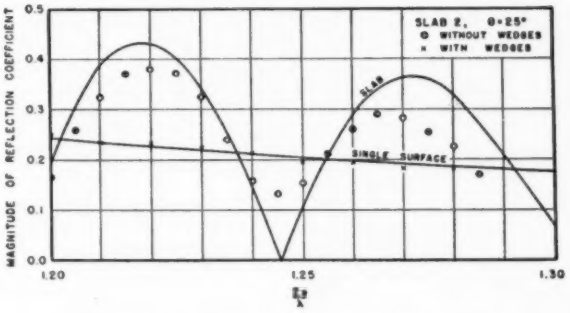


FIG. 9. The efficiency of the matching wedges.

SLAB MEASUREMENTS

The theory yields the transmission and reflection coefficients for an interface between air and a semi-infinite region composed of parallel metal plates. These coefficients apply to a single surface. In practice such a condition is difficult to obtain. One method involves the use of absorbing wedges as described in preceding sections. However, while these wedges appear to be quite effective when used with metal plate media composed of square tubes, satisfactory results could not be obtained with parallel plate structures. For this reason it was decided to make use of Lengyel's theoretical work on slabs, i.e., parallel plate structures of finite depth bounded on both sides by air.

Let r_1 and t_1 represent respectively the reflection and transmission coefficients for a wave traveling from air into a metal-plate medium and r_2 and t_2 the corresponding coefficients for a wave traveling from

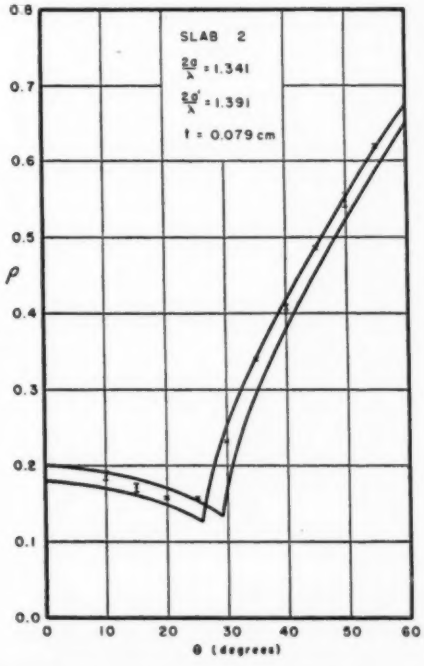


FIG. 10. Amplitude *vs* angle showing probable error.

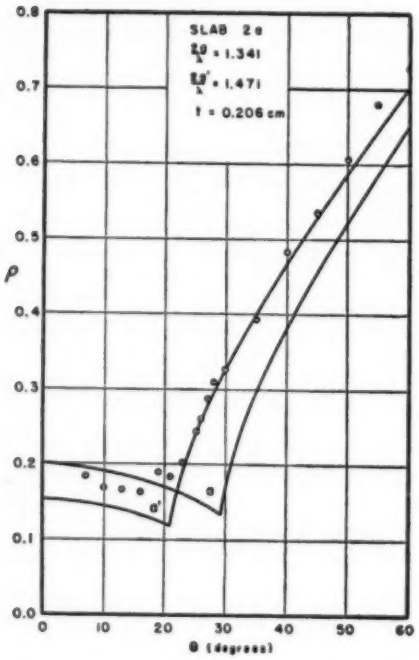


FIG. 11(a). The effect of plate thickness.

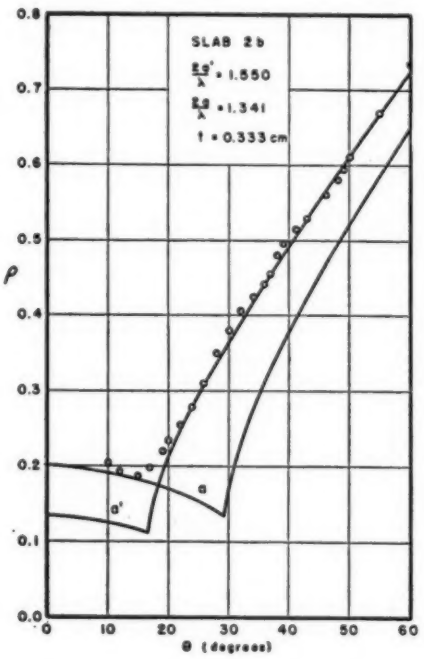


FIG. 11(b). The effect of plate thickness.

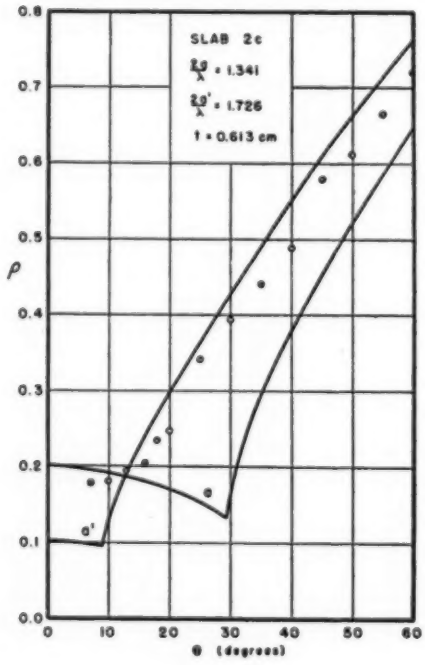


FIG. 11(c). The effect of plate thickness.

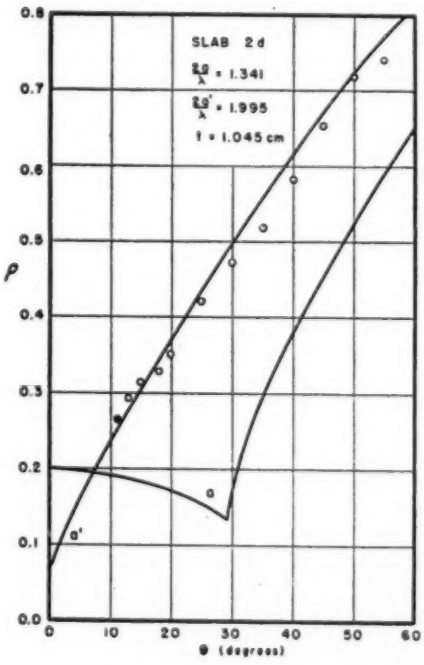


FIG. 11(d) The effect of plate thickness.

the metal-plate medium into air. Lengyel finds then that

$$|r_1| = |r_2|, \quad \mathrm{Arg}t_1 = \mathrm{Arg}t_2,$$

and

$$\mathrm{Arg}r_1 + \mathrm{Arg}r_2 - 2\mathrm{Arg}t_1 = \pm\pi.$$

These relations serve to define approximately the properties of a slab in terms of the single surface coefficients of Carlson and Heins.

The reflection coefficient of a slab is

$$R = r_1 + \frac{r_2 t_1 t_2 \exp(-2j\delta_2)}{1 - r_2^2 \exp(-2j\delta^2)},$$

where δ_2 is the phase delay along the thickness of the slab. Lengyel uses the notation:

$$r_1 = \epsilon^{-j\rho'}, \quad r_2 = \epsilon^{-j\rho''}, \quad t_1 = q\tau\epsilon^{-j\tau'}, \quad t_2 = q^{-1}\tau\epsilon^{-j\tau'}.$$

In terms of these quantities the reflection coefficient of a slab is given by $R=|R|e^{-iR'}$, where

$$|R|^2=\frac{1}{[(1/\rho)-\rho/\sin\psi]^2+1},$$

$$\mathrm{Arg}R=-R'=\tan^{-1}\frac{\cos(\psi+\rho')-\rho^2\cos(\psi-\rho')}{\sin(\psi+\rho')+\rho^2\sin(\psi-\rho')},$$

where $\psi=\rho''+\delta_2$. The above expressions for the phase and amplitude are derived in Appendix A, Tables of $-\rho'/2\pi$. Lengyel gives tables of ρ, ρ', and ρ''. It is important to note that these equations leave an uncertainty in the sign of R which amounts to an uncertainty in Arg$R$ of π. Physically this is connected with the fact that Arg R changes by π each time $|R|$ passes through zero. The sense of this phase shift is such that as $2a/\lambda$ is increased, Arg$R$ increases by π each time $|R|$ passes through zero.

In the ideal case, treated theoretically, the sense of this phase shift is of little importance since it leads to an uncertainty of 2π in the phase of R which is meaningless. However, in practice, the magnitude of R does not pass through zero, for a number of reasons, so that the phase shift as $|R|$ passes through a minimum is less than π and it is necessary to ascribe the correct sense to the shift.

The zeros of $|R|$ correspond to the zeros of $\sin\psi$ so that in order to determine the phase of the reflection coefficient for a slab, it is necessary to consider the number of zeros of $\sin\psi$ occurring between some reference wavelength and the wavelength under consideration. A convenient reference point is that for which $2a/\lambda=1$. Under these conditions Arg$R=-\rho'$ which is equal to approximately $\pi/4$. Furthermore $\psi=\pi$. As $2a/\lambda$ increases, ψ increases. From this information the number of zeros of $\sin\psi$ lying between $2a/\lambda=1$ and some greater value can be easily determined and the required number of π added to give the correct sense to R.

The value $\psi=\pi$ which corresponds to $2a/\lambda=1$ is not a zero of $|R|$ since $|R|$ assumes an indeterminate form

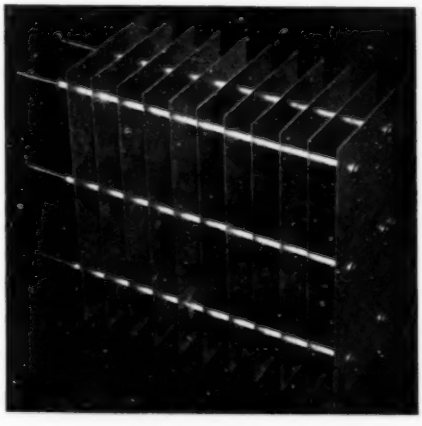


FIG. 12. Slab 3.

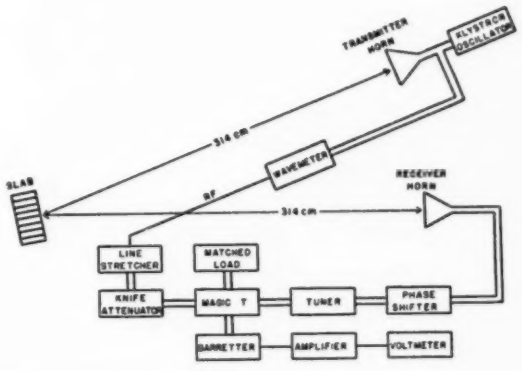


FIG. 13. Diagram of apparatus for phase measurements.

under these conditions. A limiting process shows that $|R|$ approaches a finite value of slightly less than 1 for all finite depths of slab as $2a/\lambda$ approaches 1.

The predictions of the theory were tested using Slab 3 consisting of 11 plates spaced 1.920 cm. The plate spacing was maintained within 0.008 cm with spacing tubes of 0.48 cm-diameter. The other dimensions are given in Table I. Slab 3 is shown in Fig. 12.

The source of radiation used in this series of experiments was a Varian X-13 klystron, tunable over the range 2.4 cm to 3.7 cm. A regulated power supply was constructed to operate this tube. The measurements were made using the spectrometer-like device described earlier. The transmitting, receiving, and slab apertures were all approximately equal, and the distances transmitter to slab and slab to receiver were both 314 cm. A large electromagnetic horn was used as the transmitter antenna. Phase measurements were made using a 'magic T' mixer in conjunction with a squeeze-section phase-shifter.[5] A schematic diagram of the apparatus used for phase measurements is shown in Fig. 13.

Upon selection of the desired frequency from the klystron, the following procedure was used for making both amplitude and phase measurements on Slab 3. The normalizing metal plate was placed on the front face of the slab. The RF cable between the line stretcher and wave meter which brings in the reference signal was disconnected. The tuner was adjusted until a low standing wave ratio was attained as indicated by the movement of the dial of the phase-shifter. The voltmeter reading was recorded, the normalizing plate was removed, and the voltmeter reading for the slab was obtained. This provided amplitude information. The reference signal was then brought in and adjusted by means of the knife attenuator so that its magnitude would produce the same standing wave ratio between normalizing plate and reference signal and between slab and reference signal. The phase shifter was adjusted for minimum voltmeter reading and the micrometer dial reading recorded. The normalizing plate was again

[5] A description of this phase-shifter will appear in another journal.

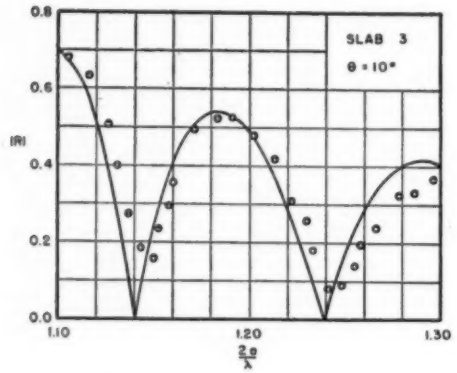


FIG. 14(a). Amplitude of reflection coefficient for Slab 3.

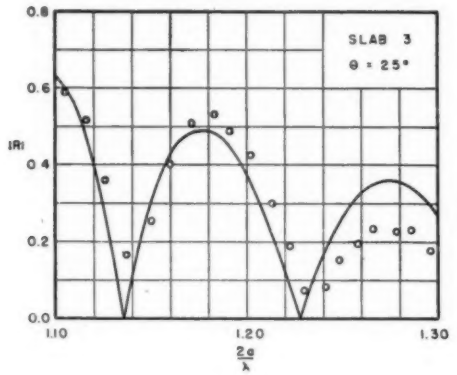


FIG. 14(b). Amplitude of reflection coefficient for Slab 3.

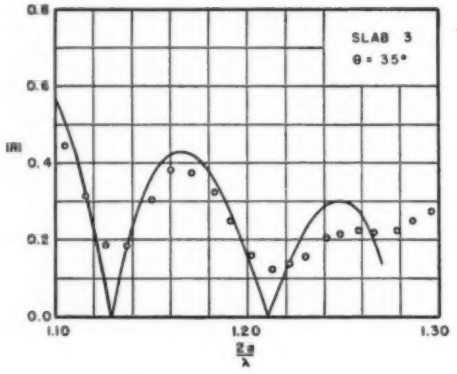


FIG. 14(c). Amplitude of reflection coefficient for Slab 3.

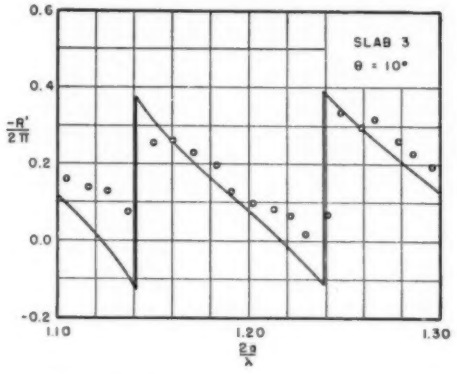


FIG. 14(d). Phase of reflection coefficient for Slab 3.

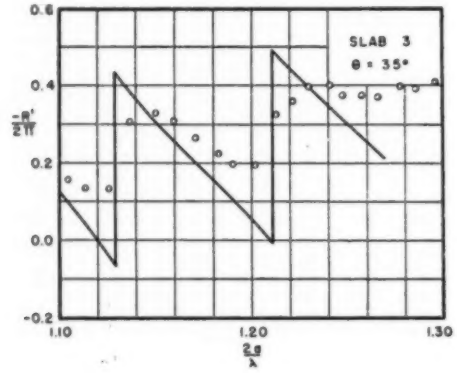


FIG. 14(e). Phase of reflection coefficient for Slab 3.

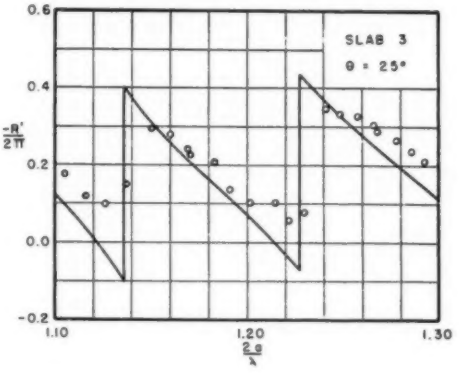


FIG. 14(f). Phase of reflection coefficient for Slab 3.

placed in position, the phase shifter adjusted for a minimum and the reading recorded. From the calibration of the phase shifter the phase shift was then determined.

Data were taken for a range of $2a/\lambda$ from 1.10 to 1.30 and for angles of incidence of 10°, 25°, and 35°, and are shown in Fig. 14. The amplitude data follow the predictions of the theory quite well. However, there is a systematic error which amounts to a displacement of the experimental curve in the direction of increased $2a/\lambda$. The agreement of the measured phase is not so good although the same systematic error is evident. In addition, the phase departs rather far from the theoretical predictions in the vicinity of the regions where $|R|$ attains its minimum values. These discrepancies are not too surprising and are probably the result of a combination of the following effects.

1. The finite distances and apertures involved will lead to phase and angle errors whose effect is difficult to take into consideration. The apertures and distances used were chosen in order to minimize their effect, but the residual effect is hard to estimate.

2. Losses within the slab will prevent $|R|$ from actually becoming zero. If $|R|$ does not pass through zero, the phase shift will be more gradual than predicted theoretically, and the total change will be less than π.

Actually the attenuation within the slab is very small (~0.06 db/meter) and this effect should be negligible.

3. The effect of plate thickness has not been treated theoretically. However, Epstein[6] has achieved some success by considering the effect to be similar to an inductive grating shunting the slab. The result of his approximation is the addition of a small angle to the phase of r_1 which has the effect of shifting the axis of the curves to the right. This interpretation qualitatively explains the systematic displacement of the experimental points to the right on both the magnitude and phase measurements. Quantitative application involves the use of an arbitrary "scale" factor so that it is not attempted here.

Figure 15 giving the reflection coefficient for $\theta=10°$ plotted in the complex plane clearly shows that experimentally $|R|$ does not pass through zero and consequently that the phase shift is less than π. The circles are experimental points and the solid curve is theoretical. $|R|$ for the theory passes through zero, and this diagram shows the phase shift of π when this takes place.

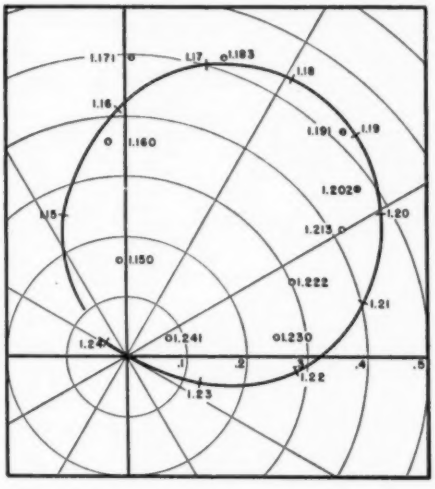


FIG. 15. Reflection coefficient in the complex plane for Slab 3 at 10° angle of incidence. Solid line curve shows the real part of R (theoretical value) as the abscissa and the imaginary part as the ordinate for values of $2a/\lambda$ from 1.24 to 1.15. The experimental points are shown as circles labeled with their proper $2a/\lambda$ value.

APPENDIX A: TABLES OF $-\rho/2\pi$

The reflection coefficient of a slab of metal-plate medium surrounded by air is given by expression (11)[7]

$$R=r_1+\frac{r_2 l_1 l_2 \exp(-2i\delta_2)}{1-r_2^2\exp(-2i\delta_2)}.$$

By using Eq. (8)[7] we obtain

$$R=\rho\exp(-i\rho')+\frac{\rho\exp(-i\rho'')\tau^2\exp(-2i\tau')\exp(-2i\delta_2)}{1-\rho^2\exp[-2i(\rho''+\delta_2)]},$$

where $\delta_2=2\pi nd/\lambda$. Letting $\psi=\rho''+\delta_2$, we get

$$R=\rho\exp(-i\rho')+\frac{\rho(1-\rho^2)\exp(\rho''+2\tau'+2\delta_2)(-i)}{1-\rho^2\exp(-2i\psi)},$$

From Eq. (9)[7] we see that

$$\rho''+2\tau'+2\delta_2=\rho'+2\psi\pm\pi.$$

Substituting this in the expression for R gives

$$R=\frac{\rho\exp(-i\rho')-\rho^3\exp[-i(\rho'+2\psi)]-\exp[-i(\rho'+2\psi)]+\rho^3\exp[-i(\rho'+2\psi)]}{1-\rho^2\exp(-2i\psi)}.$$

We now proceed in a straightforward manner to separate the real and the imaginary parts of R. Factoring $\exp(-i\psi)$ out of numerator and denominator, we have

$$R=\frac{2i\sin\psi\exp(-i\rho')}{\exp(i\psi)-\rho^2\exp(-i\psi)},$$

or

$$R=\frac{2\rho\sin\psi i\exp(-i\rho')[(\exp(-i\psi)-\rho^2\exp(i\psi)]}{1-\rho^2[\exp(2i\psi)+\exp(-2i\psi)]+\rho^4}.$$

Multiply and obtain

$$R=\frac{2\rho\sin\psi[i\cos(\psi+\rho')+\sin(\psi+\rho')-i\rho^2\cos(\psi-\rho')+\rho^2\sin(\psi-\rho')]}{1-2\rho^2\cos2\psi+\rho^4}.$$

We have finally the following expression for R in which real and imaginary parts are separated:

$$R=\frac{2\rho\sin\psi\{[\sin(\psi+\rho')+\rho^2\sin(\psi-\rho')]+i[\cos(\psi+\rho')-\rho^2\cos(\psi-\rho')]\}}{(1-\rho^2)^2+4\rho^2\sin^2\psi}.$$

[6] D. J. Epstein, Tech. Report 42, Lab. Insulation Research Massachusetts Institute of Technology (1950).
[7] B. A. Lengyel, J. Appl. Phys. **22**, 263 (1951). The equation numbers refer to those used in this reference.

From this expression for R, we have for the phase of R

$$\mathrm{Arg}R=\tan^{-1}\frac{\cos(\psi+\rho')-\rho^2\cos(\psi-\rho')}{\sin(\psi+\rho')+\rho^2\sin(\psi-\rho')}.$$

Also the amplitude of R is

$$|R|=\frac{2\rho|\sin\psi|\{1-\rho^2[\cos(\psi+\rho')\cos(\psi-\rho')-\sin(\psi+\rho')\sin(\psi-\rho')]+\rho^4\}^{\frac{1}{2}}}{(1-\rho^2)^2+4\rho^2\sin^2\psi}.$$

Now we note that

$$\cos[(\psi+\rho')+(\psi-\rho')]=\cos2\psi,$$

and making use of the relation for the cosine of the sum of two angles, we obtain

$$|R|=\frac{2\rho|\sin\psi|}{[(1-\rho^2)^2+4\rho^2\sin^2\psi]^{\frac{1}{2}}}.$$

Thus, in order to calculate the theoretical expression for Arg$R$ as a function of θ and $2a/\lambda$, it is necessary to know the appropriate values of ρ' and ρ''. The quantity ρ' designates the phase of the reflection coefficient at the air-slab interface for radiation incident from air, and ρ'' is the corresponding phase for radiation incident from the slab. Lengyel has tabulated ρ'' for values of $2a/\lambda$ ranging from 1.01 to 1.30 in intervals of 0.01 and for values of θ ranging from 0° to 35° in intervals of 5°. At the time the experimental work was done, Lengyel's article[2] in which he tabulates $-\rho'/2\pi$ for $2a/\lambda=1.01$ to 1.35 in 0.01 intervals and $\theta=0°$ to 25° in intervals of 5° was not yet available. Consequently a table of $-\rho'/2\pi$, which partially duplicates his table, was calculated. Since a considerable amount of time and effort went into the construction of this table, it is presented here in Table II. The values are determined to four places and the agreement with Lengyel's table is good.

TABLE II. Tables of $-\rho'/2\pi$.

$2a/\lambda$	10°	15°	20°	25°	30°	35°
1.10	0.1198	0.1251	0.1326	0.1424	0.1551	0.1714
1.11	0.1212	0.1268	0.1347	0.1450	0.1587	0.1762
1.12	0.1227	0.1285	0.1368	0.1477	0.1624	0.1812
1.13	0.1242	0.1302	0.1390	0.1505	0.1662	0.1866
1.14	0.1257	0.1320	0.1412	0.1534	0.1701	0.1921
1.15	0.1272	0.1338	0.1434	0.1564	0.1742	0.1980
1.16	0.1287	0.1356	0.1457	0.1595	0.1785	0.2044
1.17	0.1302	0.1374	0.1481	0.1627	0.1830	0.2114
1.18	0.1317	0.1393	0.1505	0.1660	0.1877	0.2190
1.19	0.1332	0.1412	0.1530	0.1694	0.1926	0.2272
1.20	0.1348	0.1431	0.1555	0.1729	0.1977	0.2361
1.21	0.1364	0.1451	0.1581	0.1765	0.2031	0.2457
1.22	0.1380	0.1471	0.1608	0.1803	0.2089	0.2561
1.23	0.1396	0.1491	0.1635	0.1842	0.2151	0.2675
1.24	0.1412	0.1512	0.1663	0.1883	0.2217	0.2804
1.25	0.1429	0.1533	0.1692	0.1926	0.2288	0.2958
1.26	0.1446	0.1555	0.1722	0.1971	0.2364	0.3187
1.27	0.1463	0.1577	0.1753	0.2018	0.2446	0.3635
1.28	0.1480	0.1600	0.1785	0.2068	0.2539	....
1.29	0.1498	0.1623	0.1818	0.2121	0.2645	....
1.30	0.1516	0.1647	0.1852	0.2177	0.2769	....

Lengyel[4] has derived the result $\rho'=2\mathrm{Arg}K_+(k\cos\theta)$. Substituting Eq. (13) in Eq. (10) and letting $w=k\cos\theta$, we obtain

$$V(k\cos\theta)=\sum_{n=1}^{\infty}\left\{\left[\left(1-\frac{ak\sin\theta}{2\pi n}\right)^2-\frac{(ak)^2}{2\pi n}\right]^{\frac{1}{2}}-\frac{iak\cos\theta}{2\pi n}\right\}\times\exp\left(\frac{iak\cos\theta}{2\pi n}+\frac{ak\sin\theta}{2\pi n}\right).$$

From this expression it follows that

$$\mathrm{Arg}V(k\cos\theta)=-\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{ak\cos\theta}{2\pi n[1-(ak\sin\theta/\pi n)]^{\frac{1}{2}}}-\frac{ak\cos\theta}{2\pi n}\right\}.$$

Similar expressions result for Arg$W(k\cos\theta)$ and Arg$U(k\cos\theta)$, so that, letting $w=k\cos\theta$, we obtain from Eq. (9)

$$\mathrm{Arg}K_+(k\cos\theta)=\frac{ak\cos\theta}{\pi}\ln2-(A_1+B_1)-\sum_{n=2}^{\infty}(A_n+B_n-C_n),$$

where

$$A_n=\sin^{-1}\frac{ak\cos\theta}{2\pi n[1-(ak\sin\theta/\pi n)]^{\frac{1}{2}}},$$

$$B_n=\sin^{-1}\frac{ak\cos\theta}{2\pi n[1+(ak\sin\theta/\pi n)]^{\frac{1}{2}}},$$

$$C_n=\sin^{-1}\frac{ak\cos\theta}{\pi n[1-(ak\sin\theta)^2/\pi n]},$$

and $ak/\pi=2a/\lambda$. We then have for $-\rho'/2\pi$,

$$-\rho'/2\pi=\frac{1}{\pi}\left[A_1+B_1-\frac{ak\cos\theta}{\pi}\ln2+\sum_{n=2}^{\infty}(A_n+B_n-C_n)\right].$$

An approximate method of computation was used. A remainder term

$$S_k=\sum_{n=6}^{k}(A_n+B_n-C_n),$$

was computed for $2a/\lambda=1.10$ and $\theta=0°$, the number of terms k being determined by the criterion that the kth term be less than 10^{-5}. This value of S_k was used in the computation for 10° through 35° also, since little error is introduced by this procedure and since the calculation is much simpler for $\theta=0°$. The procedure was repeated for $2a/\lambda=1.15$, 1.20, 1.25, and 1.30. The intermediate values of $2a/\lambda$ were then obtained by interpolation. Barring numerical errors, this method should yield four-place accuracy.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Method for Determining Magnetic Moments and for Measuring Susceptibilities and Permeabilities*

S. J. Barnett
California Institute of Technology, Pasadena, and University of California, Los Angeles, California
(Received April 28, 1952)

Fundamentally, the method is closely related to one described by W. Weber in 1852 in his *Elketrodynamische Maasbestimmungen* and applied by him to the measurement of very small susceptibilities in cases in which it was desirable to exclude eddy currents from the specimens under examination. It is capable of being modified so as to measure magnetic moments and magnetic susceptibilities, large and small, under a great variety of conditions. The method has the advantage, for low permeabilities, over that of permeameters, that it measures precisely the (small) *difference* between the permeability of the specimen and that of air, and not the *ratio* of the two. It determines the magnetic moment of the specimen and the susceptibility k, from which the permeability μ may be derived from the relation $\mu=1+4\pi k$.

THE arrangement used in my laboratory at the California Institute of Technology is illustrated in Fig. 1. An essentially uniform magnetic field with vertical intensity H is produced throughout a considerable region inside a long solenoid, with axis vertical. Within their central portions C and C' in this region are placed symmetrically, with axes vertical, two similar solenoids A and B. They are connected in circuit, in series-opposition, through a shunted galvanometer G, a rheostat R, and two coils HS and N, whose uses will be indicated later. The coils A and B are so constructed that both together, if traversed by unit current, would produce an essentially uniform field of strength F throughout each of the regions where the specimen S or a calibrated coil TC, is located when at rest. That is, the *coil constant* at either C or C' is F.

If the specimen S, with volume V and susceptibility k, is placed at C, it will acquire a magnetic moment $M=kVH$. (See below for effect, in some cases, of demagnetizing factor.) The magnetic flux which the specimen now sends through the galvanometer (or secondary) circuit can be shown to be

$$\Phi_s=FM\equiv FkVH.\dagger$$

If now the specimen is quickly moved from C to C', in the uniform field of strength H, no change will occur in its moment, but the flux will change from Φ_s to $-\Phi_s$, so that the total change will be $\Delta\Phi_s=2\Phi_s=2FM=2FkVH$. This will produce a galvanometer throw $D_s=A\Delta\Phi_s=A2FM=A2FkVH$, where A is a constant.

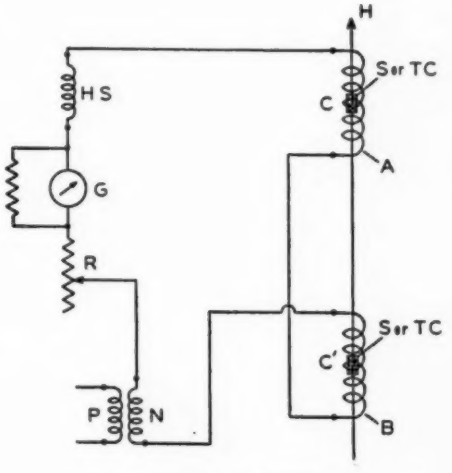


FIG. 1. Diagram of apparatus.

* Near the end of World War II the local fabricators of the B-29, in order to expedite its production, found it necessary to have certain magnetic tests made in this region instead of in the East; and I was asked to undertake the work. This paper was written at the request of the Army for a description of the process to be employed. Although this process is very simple and very useful, it is not well known, so that I have recently decided, on the advice of an experienced colleague, to publish it here. See also A. Piccard, Gen. Arch. 44, 467 (1917), and P. Weiss and A. Piccard, Gen. Arch. 45, 329 (1918).

† A body S with magnetic moment M placed in a coil C in a region within which the coil constant is uniform and equal to F sends through the coil the flux $\phi_s=FM$. This formula can be found in Mascort and Jouhert. A quite simple and quite different demonstration follows.

Replace S by an equivalent coil C' with such an area a and traversed by such a current I that its moment aI is equal to M. Let a current I traverse C. Then the flux ϕ_c sent by C through C' is $\phi_c=FIa\cos A$, where A is the angle between the normal to a and the magnetic intensity FI. The flux sent by C through C', *viz.*, $\phi_c=FIa\cos A$, is equal to the flux sent through C by either C' or S, *viz.*, ϕ_s. Thus $\phi_s=\phi_c=FIa\cos A=FM\cos A$. Usually, in practice, $A=0$, so that $\phi_s=FM$.

If the specimen is moved from C' to C, the throw will be reversed.

In order to calibrate the circuit we may proceed in either of the following two ways.

I. The specimen S may be replaced by a permanent magnet or an electric coil, with known moment m, and the motions repeated. The throw now produced by the motion between C and C' is

$$D_m = 2AFm.$$

Thus we have

$$\frac{D_s}{D_m} = \frac{2AFM}{2AFm} = \frac{M}{m} = \frac{kVH}{m}$$

and

$$k = (m/VH) \cdot (D_s/Dm).$$

Instead of a permanent magnet, with moment already known, we may, as in this work, use a short solenoid TC with axis vertical and traversed by a known electric current J. If the solenoid contains n turns and the mean area of each turn is a, the moment m is

$$m = Ja.$$

Thus we have

$$k = (Ja/VH) \cdot (D_s/Dm).$$

In many cases we may, instead of moving the specimen between C and C', simply reverse the current producing the field strength H (or the current J) while the specimen S or coil TC is left at C or C'.

In this method of procedure (I) a knowledge of F is required.

II. Instead of calibrating by means of a known magnetic moment m, we may calibrate by means of a known magnetic flux Φ_{HS}, such as that produced by a Hibbert magnetic standard.

In the figure, HS is one of the coils of a Hibbert Standard. If the coil cuts the flux Φ_{HS} when moved across the gap, a throw D_{HS} will be produced such that

$$D_{HS} = A\Phi_{HS}.$$

Thus we have

$$\frac{D_S \equiv 2AFkVH}{D_{HS} \equiv A\Phi_{HS}} = \frac{2FVH}{\Phi_{HS}} \cdot k$$

or

$$k = \frac{D_S}{D_{HS}} \cdot \frac{\Phi_{HS}}{2FVH}.$$

Thus in this method of calibration (II) a knowledge of F is necessary.

Both methods are used in this laboratory. All the instruments are standardized in the usual ways.

The coil N in the galvanometer circuit forms together with another coil P in the (primary) circuit, which produces the field of intensity H, a mutual inductance compensator. By moving N with regard to P the mutual inductance between the two circuits can be made zero, so that no deflection results when the primary circuit is reversed and the specimen is not present.

In experiments made by motion of the specimen, displacements from C to C' and from C' to C are made in equal numbers. The results are independent of the exact vanishing of the mutual inductance and of leakage between primary and secondary circuits.

In experiments made by reversals, the test coil TC and the specimen are placed at C in part of the work, and at C' in another part. The mean result is independent of any residual mutual inductance and of leakage.

For the weaker fields, the magnetizing field is produced by a very long uniformly wound single layer solenoid, and the secondary coils, with a large constant F, are placed outside it. For the stronger fields, strength about 200 gauss, the magnetizing field is produced by my *double solenoid*,[1] and the secondary coils with a smaller constant F are placed inside it. With the present arrangements (including specimens with volume about $1 \times 1 \times 2.5$ in.^{3}), a susceptibility as low as 0.00001, or a difference of about 0.0001 between the permeability of a specimen and unity, can be detected at either 75 or 200 gauss. The sensitivity can be increased when necessary.

In what precedes, it is assumed that the field intensity inside the specimen is the same as that in the same region when the specimen is removed; that is, the demagnetizing factor is assumed to be zero. This can be done with entirely negligible error when the permeabilities are small.

When the permeabilities are greater, an approximate demagnetizing factor can be calculated, and the procedure made exact. For specimens of the shape and volume supplied for the preliminary tests, the factor is about 2. Thus for a specimen calculated to have the permeability 1.103 without considering the demagnetizing field, the permeability becomes 1.105 when the demagnetizing factor is applied.

[1] S. J. Barnett, Phil. Mag. **40**, 519 (1920), and Glazebrook's *Dictionary of Applied Physics*, Vol. 2, p. 454.

The Temperature Dependence of the Viscosity of Liquids

F. GUTMANN AND L. M. SIMMONS
The New South Wales University of Technology, Sydney, Australia
(Received March 24, 1952)

The equation $\log\eta = A + B/(T+C)$ is obtained as a simple modification of Andrade's viscosity equation by treating the activation energy as a function of temperature, A, B, and C being free parameters. This is shown statistically to express adequately the temperature dependence of viscosity over substantial ranges of temperature for all the liquids examined, *viz.*, n-heptane; n-nonane; n-undecane; n-tridecane; n-heptadecane; benzene; viscous mineral oil; ethanol; 1,2-dibromopropane; water; mercury; molten gallium; potassium; sodium; sodium nitrate; silver bromide; lead chloride; bromine; 98–100 percent sulfuric acid.

RECENTLY an attempt was made to demonstrate (1) that even the best of the relatively simple viscosity temperature functions is inadequate to reproduce the measured values accurately over extended ranges of temperature, and (2) that relationships that reproduce the measured values with satisfactory validity are necessarily complicated and not likely to hold promise of fundamental significance.[1] We therefore draw attention to the equation

$$\log\eta = A + B/(T+C), \tag{1}$$

which expresses viscosity (η) adequately over extended ranges of temperature (T°K). A, B, and C are free parameters determined by least-square analysis of experimental data, the normal equations being

$$\begin{aligned} A\Sigma T^2 - C\Sigma\rho T + \lambda\Sigma T - \Sigma\rho T^2 &= 0, \\ A\Sigma\rho T - C\Sigma\rho^2 + \lambda\Sigma\rho - \Sigma\rho^2 T &= 0, \\ A\Sigma T - C\Sigma\rho + n\lambda - \Sigma\rho T &= 0, \end{aligned} \tag{2}$$

where $\rho = \log\eta$ and $\lambda = B + AC$.

In Table I we have compared the closeness of fit of this relatively simple equation with that of the far less simple equation proposed by Doolittle:

$$\ln\eta = -A_1 e^{500/T}\{-Ei(-500/T)\} - 4.66 + C_1 e^{500/T}, \tag{3}$$

where $Ei(-500/T)$ is the logarithmic integral (to be read from tables) and C_1 is the molecular weight dependent constant of integration. For this comparison we have used those viscosity data quoted by Doolittle which comprise sufficient pairs of readings for the statistical examination of a three parameter equation, *viz.*, four or more. The Gaussian criterion of closeness of fit has been applied in the form

$$\Omega = \frac{\Sigma(1000\eta_{obs} - 1000\eta_{calc})^2}{n - m}, \tag{4}$$

where n is the number of observed pairs and m is the number of free parameters.[2] Equation (1) has three free parameters, A, B, and C, while Eq. (3) involves two quantities, C_1 and A_1, whose numerical values were calculated from the experimental temperature-viscosity data. They are therefore statistically free parameters. It is seen from Table I that in no case is Eq. (3) significantly preferable to Eq. (1).

In regard to the fundamental significance of Eq. (1), we have derived it from the Boltzmann-Maxwell distribution law.[3] Furthermore, although Doolittle claims that "attempts to express the constant (E) in the Andrade equation as a function of temperature were particularly frustrating," we give the following alternative derivation of Eq. (1):

Harrap and Heymann[4] have pointed out that the activation energy (E) of viscosity in the equation

$$\eta = A' e^{E/kT} \tag{5}$$

for a number of compounds decreases with increasing temperature. Furthermore, Ward[5] has pointed out that E depends on the coordination of the molecules and ions in the liquid and will in most cases vary considerably with temperature. Let this variation be expressed by

$$E = E_0 + f(T), \tag{6}$$

where E_0 is independent of temperature, and $f(T)$ is a convergent power series in T.

By virtue of Bürmann's theorem[6] Eq. (6) may be written as follows:

$$E = E_0/\phi(1/T). \tag{7}$$

TABLE I. Comparison between the closeness of fit of Doolittle's equation and that of $\log\eta = A + B/(T+C)$.

Liquid	Temperature range °K	Gaussian criterion of closeness of fit for: Doolittle's equation	$\log\eta = A + B/(T+C)$
n-Heptane	263–373	49×10^{-5}	2.5×10^{-5}
n-Nonane	263–423	1.3×10^{-3}	1.6×10^{-3}
n-Undecane	263–473	1.8×10^{-2}	1.5×10^{-2}
n-Tridecane	263–473	1.21	0.18
n-Heptadecane	293–573	0.64	0.47

[1] A. K. Doolittle, J. Appl. Phys. **22**, 1031 (1951).
[2] A. G. Worthing and J. Geffner, *Treatment of Experimental Data* (John Wiley & Sons, Inc., New York, 1943), p. 261.
[3] F. Gutmann and L. M. Simmons, Aust. J. Sci. **13**, 109 (1951).
[4] B. S. Harrap and E. Heymann, Chem. Revs. **48**, 45 (1951).
[5] A. G. Ward, Trans. Faraday Soc. **33**, 88 (1937).
[6] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, London, England, 1935), pp. 128 *ff*.

TABLE II. The fit of $\log\eta = A + B/(T+C)$.

Liquid	Temperature range °K	*SPE*	*n*	*A*	*B*	*C*	Re
n-Heptane	263–373	0.046	4	−3.7278	401.35	−19.6	a
n-Nonane	263–423	0.42	5	−3.78256	448.65	−19.8	a
n-Undecane	263–473	1.5	6	−3.75431	461.47	−41.8	a
n-Tridecane	263–473	1.4	6	−3.62961	419.56	−73.89	a
n-Heptadecane	293–573	1.8	7	−3.84518	616.99	−42.11	a
Benzene	273–343	0.37	8	−3.65993	332.4	−67.4	b
Ethanol	273–343	0.076	8	−5.90519	1922.5	+189.7	b
1,2-Dibromopropane	273–413	0.44	15	−3.3929	380.52	−56.30	b
Water	264–373	0.62	27	−3.540	222.21	−149.37	c
Mercury	253–613	0.034	18	−2.21377	101.66	−42.28	d
Gallium	303–1373	1.6	10	−2.46363	337.73	+131.16	e
Potassium	340–625	0.88	10	−3.00893	251.05	+6.48	f
Sodium	393–628	0.27	10	−3.11229	404.62	+54.99	f
Bromine	270–305	0.12	5	−2.50041	60.96	−170.63	g
Sodium nitrate	581–691	0.20	12	−3.43407	1749.79	+340.16	g
Silver bromide	882–1076	1.5	5	−2.36919	284.11	−437.33	g
Lead chloride	771–881	0.59	12	−1.85141	76.81	−643.24	g
98–100 percent H_2SO_4	273–363	4.0	11	−4.80671	1257.24	+5.27	h
Viscous mineral oil	293–368	0.51	16	−3.317	558.35	−169.03	i

[a] See reference 1.
[b] E. W. Madge, J. Phys. Chem. **34**, 1605 (1930).
[c] D. H. Black, Nature **125**, 583 (1930).
[d] *Handbook of Chemistry and Physics* (1948), thirtieth edition, p. 1733.
[e] C. J. Smithells, *Metals Reference Book* (Butterworth Scientific Publications, London, 1949), p. 415.
[f] Y. S. Chiong, Proc. Roy. Soc. (London) **A157**, 272 (1936).
[g] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1928), **7**, 212.
[h] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1928), **5**, 11.
[i] See reference 7.

This inversion is subject to the usual limitations that $f(T)$ and all its derivatives are analytic and holomorphic within a given region of temperature. $\phi(1/T)$, whatever may be its form, may then be expanded as a power series in $1/T$ thus:

$$\phi(1/T) = 1 + a + bT^{-1} + cT^{-2} + \cdots, \tag{8}$$

where one, some, or all of the coefficients $a, b, c \cdots$ may vanish.

Should all the coefficients vanish, then $\phi(1/T) = 1$ and $E = E_0$; if all but a should vanish, then $E = E_0/(1+a)$. In each of these cases E becomes independent of temperature, which is the conventional form. A closer approximation is achieved by retaining the linear term in $1/T$, then

$$\phi(1/T) \doteqdot 1 + a + b/T, \tag{9}$$

leading to

$$E = E_0/(1+a+b/T). \tag{10}$$

Substituting C for $b/(1+a)$ and E' for $E_0/(1+a)$ in Eqs. (5) and (10), we obtain

$$\eta = A' e^{E'/k(T+C)}, \tag{11}$$

which is the exponential form of Eq. (1).

Table II shows that Eq. (1) is quite adequate in representing temperature-viscosity data over substantial ranges in temperature for representatives of widely differing classes of liquids, including ionic, metallic, and molecular liquids. In this table, *SPE* stands for the standard percentage error of fit and is defined by

$$SPE = 100\left[\frac{\Sigma(\eta_{\text{calc}} - \eta_{\text{obs}}/\eta_{\text{obs}})^2}{n}\right]^{\frac{1}{2}}. \tag{12}$$

For all the pure liquids listed in Table II the *SPE*'s lie within the experimental errors; furthermore, the possible applicability of Eq. (1) to mixed liquids is demonstrated by the case for the viscous mineral oil investigated by Black.[7] An empirical equation of similar form has also been shown to express adequately the viscosities of some tractor oils,[8] some molten glasses,[9] and some supercooled liquids.[10]

[7] D. H. Black, Nature **125**, 581 (1930).
[8] B. Deryagin and M. Kuskov, Neftyanoe Khozyaistvo **26**, 55 (1934).
[9] G. S. Fulcher, J. Am. Ceram. Soc. **8**, 339 (1925).
[10] A. Tammann, Z. anorg. u. allgem. Chem. **156**, 245 (1926).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

On the Calorimetric Heat Exchange Correction

H. D. VASILEFF
Sylvania Electric Products Incorporated, Physics Laboratories, Bayside, New York
(Received March 24, 1952)

A method has been developed for determining the low temperature heat capacity of a cylindrical sample from heating and cooling curves. The effect of temperature gradient in the sample is included.

I. INTRODUCTION

HEAT exchange between a calorimeter and its surroundings requires careful consideration in any calorimetric experiment. The measured temperature rise has to be corrected for heat exchange, and in some cases large errors can be introduced.

In the case of the low temperature calorimeter, several causes of heat exchange have to be considered. The most important of these is radiation to or from the surrounding jacket. Solid conduction can take place by way of the electrical leads and the support both to the jacket and also from the top of the cryostat, which is at room temperature. The heat exchange with the jacket due to radiation and gas conduction can be approximated by Newton's law, if the thermal head is small. The conduction of heat from the top of the cryostat can usually be treated as constant since experience shows that it causes a slow uniform increase of temperature in the calorimeter. Therefore, if T is the calorimeter temperature and T_j the jacket temperature,

$$dT/dt = R'(T_j - T) + H, \tag{1}$$

where R' is the leakage modulus of the system and H allows for the constant addition of heat.

The correction to the observed temperature rise for heat exchange is usually determined by a graphical method developed by Keesom and Kok.[1] This method is open to criticism and is discussed in this paper.

In order to apply the method, the calorimeter temperature is read at frequent intervals before, during, and after the heating period. The temperature T is plotted against time t, giving a graph of the form shown by T_F, T_H, T_A in Fig. 1 or Fig. 2. In order to determine the corrected temperature rise from this graph, the following assumptions were made by Keesom and Kok.

The heat capacity C' of the calorimeter and contents is constant during the period considered. The temperature T of the calorimeter is uniform throughout. The heat production in the heating period is constant (Q calories per sec). The temperature T_B of the surrounding bath is some regular function of time $f(t)$. If no heating occurred, the temperature of the calorimeter would be $T_F = g(t)$. (This coincides with the real temperature of the calorimeter in the fore-period $t<0$.) The real temperature in the heating period is T_H and in the after-period T_A. The heat exchange with the immediate surroundings is proportional to the temperature difference as in Eq. (1) in which T_j can be identified with T_F which is not necessarily equal to T_B.

Under these assumptions, particularly the second, the temperature-time graph has the form shown in Fig. 1. Keesom and Kok derive for it the following equations.

During the heating period

$$(T_H - T_F) = (Q/R')\{1 - \exp(-R't/C')\}. \tag{2}$$

During the after-period

$$(T_A - T_F) = b \exp\{-R'(t - t_1)/C'\}, \tag{3}$$

where t_1 is the end of the heating period and

$$b = (Q/R')\{1 - \exp(-R't_1/C')\}$$

is the observed temperature difference at time t_1.

In the foregoing discussion no account has been taken of the possible lag in the temperature measurements. The heat exchange between the calorimeter and the jacket is primarily determined by the temperatures of the outside surface of the former and the inside surface of the latter. Error will be introduced in the heat exchange correction if the temperature of the surface of the calorimeter, during a period when its temperature is rapidly changing, is inferred from the reading of a thermometer located in its interior. When the thermometer is in the center of the calorimeter and the heating element is on the surface, the peak of the heating curve does not occur exactly at the end of the heating interval

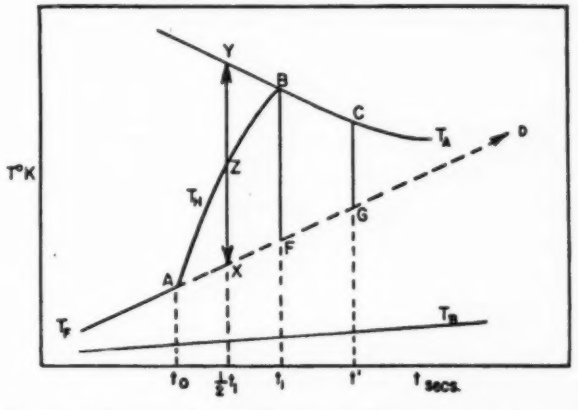


FIG. 1. The Keesom and Kok (see reference 1) heating (AB) and cooling (BC) curve. The maximum temperature occurs at time t_1, the end of the heating interval.

[1] W. H. Keesom and J. A. Kok, Leiden Comm. 219(c).

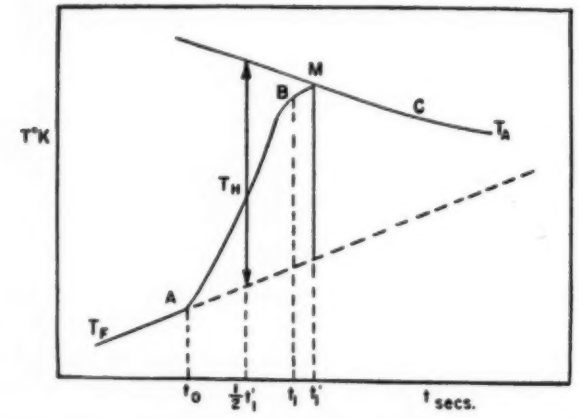


FIG. 2. The heating (AB) and cooling (BMC) curve of a cylindrical sample, having the thermometer at the center and the heat source on the surface. The end of the heating interval occurs at time t_1 and the maximum temperature at t_1'.

t_1, but rather some seconds later at t_1', Fig. 2. During the heating, a temperature gradient is set up between the center and the surface of the calorimeter and, after the end of the heating interval, the temperature at the center continues to rise for a time until temperature uniformity is attained.

In the apparatus used by Keesom and Kok the heater and resistance thermometer were close together in the central hole. The slope of the temperature curve did, therefore, change abruptly at the beginning and end of the heating interval (as in Fig. 1), but there was no assurance that lag did not occur. Keesom and Kok neglected the possibility of lag and assumed that their method of correction could be applied to the central temperature readings. Consequently, it would be better to separate the heater and thermometer so that the lag becomes evident. One can then either allow for the lag or show that it is small enough not to introduce a significant error.

The purpose of this paper is to derive equations which will properly describe the heating and cooling curves when the thermal conductivity of the sample in the calorimeter is taken into account. When the thermal conductivity approaches infinity, it is shown that the derived equations reduce to those of Keesom and Kok, Eqs. (2) and (3).

II. THE HEATING EQUATION

The shape of the sample for which the temperature distribution is to be calculated is chosen as a section of an infinite cylinder. To measure the heat capacity of this sample a resistance thermometer is placed in a cylindrical opening through its center, parallel to its axis; a heater is wound around the circular outside surface. It is clear that during the heating interval the temperature at the surface will be higher than that at the center.

The following quantities define the cylinder: r_0=radius, l=length, δ=density, c=specific heat, k=thermal conductivity. The equation of heat conduction is

$$\frac{\partial^2 U(r,t)}{\partial r^2}+\frac{1}{r}\frac{\partial U(r,t)}{\partial r}-\frac{1}{D}\frac{\partial U(r,t)}{\partial t}=0,$$

where $D=k/\delta c$ is the diffusivity of temperature.

Putting $r=\rho r_0$ so that $0\leq\rho\leq 1$, and $\tau=Dt/r_0^2$, we obtain

$$\frac{\partial^2 U(\rho,\tau)}{\partial \rho^2}+\frac{1}{\rho}\frac{\partial U(\rho,\tau)}{\partial \rho}-\frac{\partial U(\rho,\tau)}{\partial \tau}=0. \tag{4}$$

We now let $U(\rho,\tau)$ be the difference between the temperature of the cylinder and the temperature of the surrounding medium. Let heat be supplied at the surface of a rate of q calories per sec per cm^2, commencing at time $t=0$. Then the total quantity of heat supplied per sec is $Q=q2\pi r_0 l$ cal$-$sec^{-1}. We assume also that at $t=0$ the temperature is uniform and equal to that of the surroundings.

If the cylinder radiates heat according to Newton's law of cooling, the boundary conditions are

$$U(\rho,0)=0$$

$$2\pi r_0 k\frac{\partial U(r,t)}{\partial r}\bigg|_{r=r_0}=q\cdot 2\pi r_0-RU(r_0,t), \tag{5}$$

or

$$2\pi\frac{\partial U(\rho,\tau)}{\partial \rho}\bigg|_{\rho=1}=q\cdot(2\pi r_0/k)-(R/k)U(1,\tau), \tag{6}$$

where R is proportional to Newton's radiation constant so that RU is the radiation loss per unit length.

Taking now the Laplace transforms of Eqs. (4) and (6), we have

$$\frac{\partial^2 u(\rho,s)}{\partial \rho^2}+\frac{1}{\rho}\frac{\partial u(\rho,s)}{\partial \rho}-su(\rho,s)=0 \tag{7}$$

and

$$2\pi\frac{\partial u(\rho,s)}{\partial \rho}\bigg|_{\rho=1}=(q_1/s)-(R/k)u(1,s), \tag{8}$$

where $q_1=2\pi q r_0/k=Q/lk$ and s is a new parameter defined by

$$u(\rho,s)=\int_0^\infty U(\rho,\tau)\exp(-s\tau)d\tau.$$

Equation (7) is Bessel's equation and a solution finite at $\rho=0$ is $u(\rho,s)=CJ_0[i(s)^{\frac{1}{2}}\rho]$, where J_0 is Bessel's function of the first kind and C is a constant. To find the value of the constant C, substitute into Eq. (8) and

$$C=(q_1/s)\{2\pi i(s)^{\frac{1}{2}}\cdot J_0'[i(s)^{\frac{1}{2}}]+(R/k)J_0[i(s)^{\frac{1}{2}}]\}^{-1},$$

and hence,

$$u(\rho, s)=q_1J_0[i(s)^{\frac{1}{2}}\cdot\rho]/ 2\pi s\{(R/2\pi k)J_0[i(s)^{\frac{1}{2}}]+i\sqrt{s}J_0'[i(s)^{\frac{1}{2}}]\}. \quad (9)$$

By the inversion theorem, the inverse transform is

$$U(\rho, \tau)=\frac{-1}{2\pi i}\lim_{\beta\to\infty}\int_{\gamma-i\beta}^{\gamma+i\beta} \times\frac{q_1J_0[i(s)^{\frac{1}{2}}\cdot\rho]\exp(s\tau)ds}{2\pi s\{i(s)^{\frac{1}{2}}\cdot J_1[i(s)^{\frac{1}{2}}]-(R/2\pi k)J_0[i(s)^{\frac{1}{2}}]\}},$$

where the singularities of the integrand are all to the left of the real constant γ in the complex plane.

It may be shown that the integrand vanishes sufficiently rapidly at infinity for Cauchy's theorem to be applied and that $U(\rho, \tau)$ is equal to the sum of the residues of the integrand, whose singularities occur at $s=0$ and at the roots of the equation

$$i(s)^{\frac{1}{2}}J_1[i(s)^{\frac{1}{2}}]-(R/2\pi k)J_0[i(s)^{\frac{1}{2}}]=0.$$

If the sequence of roots are $\alpha_1, \alpha_2\cdots$ then

$$\alpha_nJ_1(\alpha_n)-(R/2\pi k)J_0(\alpha_n)=0. \quad (10)$$

where $\alpha_n=i(s)^{\frac{1}{2}}$ are tabulated in Carslaw and Jaeger.[2] The residue at $s=0$ is $q_1k/R=Q/lR$, and the residue at $s=-\alpha_n^2$ for $n=1, 2, 3\cdots$ is the sum

$$\sum_{n=1}^{\infty}-q_1J_0(\rho\alpha_n)\exp(-\alpha_n^2\tau)/2\pi\alpha_n^2 \times\{(J_0(\alpha_n)/2\alpha_n^2)(\alpha_n^2+R^2/4\pi^2k^2)\}.$$

Therefore, since $r=\rho r_0$, we obtain

$$U(r, t)=\frac{Q}{lR}[1-\sum_{n=1}^{\infty}RJ_0(r\alpha_n/r_0) \times\exp(-D\alpha_n^2t/r_0^2)/\pi kJ_0(\alpha_n)\{\alpha_n^2+R^2/4\pi^2k^2\}]. \quad (11)$$

Hence, the temperature at the center of the cylinder is given by

$$U(0, t)=\frac{Q}{lR}[1-\sum_{n=1}^{\infty}R\exp(-D\alpha_n^2t/r_0^2)/\pi kJ_0(\alpha_n) \times\{\alpha_n^2+R^2/4\pi^2k^2\}] \quad (12)$$
$$=\frac{Q}{lR}[1-\sum_{n=1}^{\infty}f_n].$$

It can easily be shown that this equation reduces to that of Keesom and Kok, if the thermal conductivity k is allowed to become infinite. Since all the roots α_n for $n>1$ approach a finite value ($\neq 0$) as $k\to\infty$ then $f_n\to 0$ for $n>1$, and

$$\lim_{k\to\infty}f_1=\exp(-Rt/\pi\delta cr_0^2).$$

[2] Carslaw and Jaeger, *Heat Conduction in Solids* (Oxford University Press, New York, 1947).

Hence,

$$U(0, t)=U(r_0, t)=(Q/R')\{1-\exp(-R't/C')\}, \quad (13)$$

where $R'=lR$ and $C'=l\pi r_0^2\delta c$ is the heat capacity of the cylinger. Equation (13) corresponds exactly to that of Keesom and Kok, Eq. (2). The "heating equation," Eq. (12), is later joined to the "cooling equation" derived in the next section.

III. THE COOLING EQUATION

Writing the equation of heat conduction in the form

$$\frac{\partial^2U}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial U}{\partial\rho}-\frac{r_0^2}{D}\frac{\partial U}{\partial t}=0 \quad (14)$$

and putting $U=V\exp(-D\alpha_n^2t/r_0^2)$, we obtain

$$\frac{\partial^2V}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial V}{\partial\rho}+\alpha_n^2V=0.$$

Hence, $V=CJ_0(\rho\alpha_n)$,

$$U(\rho, t)=CJ_0(\rho\alpha_n)\exp(-D\alpha_n^2t/r_0^2),$$
$$U(r, t)=CJ_0(r\alpha_n/r_0)\exp(-D\alpha_n^2t/r_0^2). \quad (15)$$

When there is no heating, the boundary equation becomes

$$2\pi r_0k\left.\frac{\partial U(r, t)}{\partial r}\right|_{r=r_0}=-RU(r_0, t). \quad (16)$$

Substituting Eq. (15) into Eq. (16), we obtain

$$\alpha_nJ_1(\alpha_n)-(R/2\pi k)J_0(\alpha_n)=0.$$

Since there is an infinite sequence of roots,

$$U(r, t)=\sum_{n=1}^{\infty}C_nJ_0(r\alpha_n/r_0)\exp(-D\alpha_n^2t/r_0^2). \quad (17)$$

If the heating of the cylinder had previously ceased at time $t=t_1$, then the temperature of the cylinder at the beginning of the cooling period is

$$U(\rho, t_1)=\frac{Q}{lR}[1-\sum_{n=1}^{\infty}RJ_0(\rho\alpha_n) \times\exp(-D\alpha_n^2t_1/r_0^2)/\pi kJ_0(\alpha_n)\{\alpha_n^2+R^2/4\pi^2k^2\}]. \quad (18)$$

The following integrals will be of use:

$$\int_0^1\rho J_0(\rho\alpha_n)\cdot J_0(\rho\alpha_m)d\rho=0 \quad m\neq n,$$

$$\int_0^1\rho J_0^2(\rho\alpha_m)d\rho=(1/2\alpha_m^2)(p^2+\alpha_m^2)J_0^2(\alpha_m),$$

$$\int\rho J_0(\rho\alpha_m)d\rho=(1/\alpha_m)\cdot J_1(\alpha_m),$$

where

$$\alpha_m J_1(\alpha_m) - pJ_0(\alpha_m) = 0. \quad (19)$$

Multiplying both sides of the equation

$$U(\rho, t_1) = \sum_{n=1}^{\infty} C_n J_0(\rho\alpha_n) \exp(-D\alpha_n^2 t_1/r_0^2)$$

by $\rho J_0(\rho\alpha_m)d\rho$ and then integrating, we obtain

$$C_m = \frac{Q}{lR}[\{2\alpha_m J_1(\alpha_m)/(\alpha_m^2+p^2)J_0^2(\alpha_m)\} \times \exp(D\alpha_m^2 t_1/r_0^2) - 2p/J_0(\alpha_m)\cdot(\alpha_m^2+p^2)].$$

Substituting for C_m in Eq. (17) and making use of (19), one obtains

$$U(\rho, t) = \frac{Q}{lR}\sum_{m=1}^{\infty}\{2pJ_0(\rho\alpha_m)/(\alpha_m^2+p^2)\cdot J_0(\alpha_m)\} \times\{\exp(D\alpha_m^2 t_1/r_0^2)-1\}\exp(-D\alpha_m^2 t/r_0^2). \quad (20)$$

At the center of the cylinder the temperature is given by

$$U(0, t) = \frac{Q}{lR}\sum_{m=1}^{\infty}\{2p/(\alpha_m^2+p^2)\cdot J_0(\alpha_m)\} \times\{1-\exp(-D\alpha_m^2 t_1/r_0^2)\} \times\exp\{-D\alpha_m^2(t-t_1)/r_0^2\}. \quad (21)$$

Again it can be shown that as the thermal conductivity $k\to\infty$ or as $p\to 0$

$$\lim_{p\to 0} U(0, t) = \frac{Q}{lR}\{1-\exp(-R't_1/C')\} \times\exp\{-R'(t-t_1)/C'\} = b\exp\{-R'(t-t_1)/C'\},$$

which corresponds exactly to Keesom and Kok's equation representing the cooling of a sample of infinite conductivity.

IV. CALCULATION OF THE TIME LAG AND HEAT CAPACITY

If Eq. (21) is differentiated with respect to time and equated to zero, one obtains for the time t_1' at which the maximum temperature occurs, Fig. 2,

$$\sum_{n=1}^{\infty}\frac{\alpha_n^2}{(\alpha_n^2+p^2)\cdot J_0(\alpha_n)}\{1-\exp(-D\alpha_n^2 t_1/r_0^2)\} \times\exp\{-D\alpha_n^2(t_1'-t_1)/r_0^2\} = 0, \quad (22)$$

where $p = R'/2\pi lk$. Since the α_n are roots of Eq. (10), then for small values of p the α_n are approximately equal to the roots of $\alpha_n J_1(\alpha_n) = 0$. The first four successive roots for $p=0$, as taken from Carslaw and Jaeger,[2] are $\alpha_1=0$, $\alpha_2=3.8317$, $\alpha_3=7.0156$, $\alpha_4=10.1735$. Since the successive values of α_n increase rapidly, the infinite series Eq. (22) converges very rapidly. As an approximation, it is sufficient to consider only the first two terms of the series, with the condition $\alpha_n^2/(\alpha_n^2+p^2)\simeq 1$ always. Then

$$\frac{1}{J_0(\alpha_1)}\{1-\exp(-D\alpha_1^2 t_1/r_0^2)\} \times\exp\{-D\alpha_1^2(t_1'-t_1)/r_0^2\} + \frac{1}{J_0(\alpha_2)}\{1-\exp(-D\alpha_2^2 t_1/r_0^2)\} \times\exp\{-D\alpha_2^2(t_1'-t_1)/r_0^2\} = 0. \quad (23)$$

Now, since $\alpha_2^2 >> \alpha_1^2$, $J_0(\alpha_1)\simeq 1$ and

$$\exp\{-D\alpha_2^2 t_1/r_0^2\} << 1,$$

Eq. (23) may be written as

$$\exp\{D(t_1'-t_1)\alpha_2^2/r_0^2\} = -1/J_0(\alpha_2)\{1-\exp(-D\alpha_1^2 t_1/r_0^2)\}. \quad (24)$$

From the equation $\alpha J_1(\alpha) - pJ_0(\alpha) = 0$, one can easily deduce from the series expansion that for $\alpha<1$, $\alpha_1^2\simeq 2p$ where p is very small. Hence, since $D=k/\delta c$, one obtains $D\alpha_1^2/r_0^2 = R'/C'$. Since

$$J_0(\alpha_2) = J_0(3.8317) = -0.403,$$

Eq. (24) becomes

$$t_1'-t_1 = -(C'/\pi lk\alpha_2^2) \ln\{0.403[1-\exp(-R't_1/C')]\}. \quad (25)$$

This equation does predict that the maximum temperature occurs at a time t_1', after the end of the heating period t_1, and also that when $k\to\infty$, $t_1'\to t_1$.

An expression for $t_1'-t_1$ may be found in another way. If one considers only the first two terms in the expansion for $U(0, t)$, Eq. (21), one obtains, after substituting from Eq. (24),

$$U(0, t_1') = -\{Q/R'J_0(\alpha_1)\}\left\{\frac{\alpha_1^2}{\alpha_2^2}-1\right\} \times\{1-\exp(-R't_1/C')\}\exp\{-R'(t_1'-t_1)/C'\} = (Q/R')\{1-\exp(-R't_1/C')\} \times\exp\{-R'(t_1'-t_1)/C'\}. \quad (26)$$

Similarly, the expression for $U(0, t_1)$ becomes

$$U(0, t_1) = (Q/R'J_0(\alpha_1))\{1-\exp(-R't_1/C')\} \times\left\{1-\frac{\alpha_1^2}{\alpha_2^2}\exp[D(\alpha_2^2-\alpha_1^2)(t_1'-t_1)/r_0^2]\right\}. \quad (27)$$

By taking the ratio of Eqs. (26) and (27), one obtains the result

$$U(0,t_1')/U(0,t_1)=\{\exp(-\theta)\}\Big/\left\{\exp(-\theta)-\frac{\alpha_1^2}{\alpha_2^2}\right\}, \quad (28)$$

where

$$\theta=D\alpha_2^2(t_1'-t_1)/r_0^2=(\pi lk\alpha_2^2/C')(t_1'-t_1).$$

Solving Eq. (28) for $t_1'-t_1$ gives

$$(t_1'-t_1)=(C'/\pi lk\alpha_2^2)\ln[(\alpha_2^2/\alpha_1^2)(1-U(0,t_1)/U(0,t_1'))]. \quad (29)$$

Equations (25) and (29) together result in

$$(1-U(0,t_1)/U(0,t_1'))=(\alpha_1^2/\alpha_2^2)\{0.403(1-\exp(-R't_1/C'))\}^{-1},$$

which along with Eq. (25) gives

$$\{(t_1'/t_1)-1\}\{1-U(0,t_1)/U(0,t_1')\}^{-1}=(0.403/\psi)(1-\exp\psi)\ln\{0.403(1-\exp\psi)\}, \quad (30)$$

where $\psi=-R't_1/C'$.

V. CONCLUSION

From the experimentally obtained heating and cooling curves, the quantities t_1, t_1', $U(0,t_1)$, $U(0,t_1')$, and Q are easily obtained. Equation (30) then allows for a calculation of the ratio R'/C'. Having obtained this quantity, one may then calculate R' from Eq. (26), and hence the heat capacity C'. Formulas (25) and (30) have been checked on heat capacity measurements taken in the temperature range 65°K–300°K, and yielded values of C' approximately 2 percent larger than the values obtained by applying the Keesom and Kok method to heating and cooling curves of the type shown in Fig. 2.

The author expresses his gratitude to Professor H. Grayson-Smith of the University of Alberta for his interest in the problem, and to the staff of the Theoretical Section of the Sylvania Physics Laboratories, where the problem was finally completed.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

A Statistical Approach to the Space-Charge Distribution in a Cut-Off Magnetron*

GUNNAR HOK
Department of Electrical Engineering, University of Michigan, Ann Arbor, Michigan
(Received March 28, 1952)

When the anode potential of a nonoscillating magnetron is lower than the cut-off potential, the discrete electron-to-electron interaction has an effect on the space-charge distribution which would be negligible if only resulting in fluctuations about a known steady state. However the states proposed by Brillouin, Slater, and others are not steady states in this sense. The electron interaction must, therefore, be expected to produce a steady drift away from such initial states. This paper discusses qualitatively the final steady state to be expected and its dependence on the ratio of anode voltage to cut-off voltage. Because of the involved nature of the electron distribution function no attempt is made to calculate the space-charge distribution explicitly.

I. INTRODUCTION

AS an initial state for the oscillating magnetron the nonconducting (or approximately nonconducting) condition of a dc magnetron has a considerable interest and has received appreciable attention, both of a theoretical and an experimental nature. Nonetheless, a satisfactory agreement has not been reached about the shape of the electron orbits and the detailed distribution of potential and electron density in such a space charge.

The solutions presented necessarily rest on postulates and assumptions that constitute idealizations of the real conditions. So does, of course, all physical theory, but in this case well-known factors have been neglected with the justification that their effect is judged to be so small that the approximate solution resulting will be close enough to be of value. Furthermore, the mathematical difficulties of a more rigorous approach appeared prohibitive.

It is the purpose of this paper to reconsider the conventional simplifications on which these solutions are based, to discuss whether or not they introduce appreciable errors, and to investigate the feasibility of a solution from a more realistic set of assumptions. Since the space charge is essentially a dilute gas formed by discrete electrons, classical statistical mechanics offers a logical approach to a fresh study of the problem.

We shall first place the problem into the framework of statistical mechanics. The next step is to consider the solutions proposed so far in this light and to show that they cannot possibly represent steady states. This discussion does not directly indicate whether the errors in the distribution of space charge and potential are large or small. Finally, the problem of finding a solution consistent with the laws of statistical mechanics will be investigated and the result compared with the previously suggested solutions.

* This work was undertaken as part of a magnetron research program sponsored by the U. S. Signal Corps under contract No. DA-36-039 SC-5423.

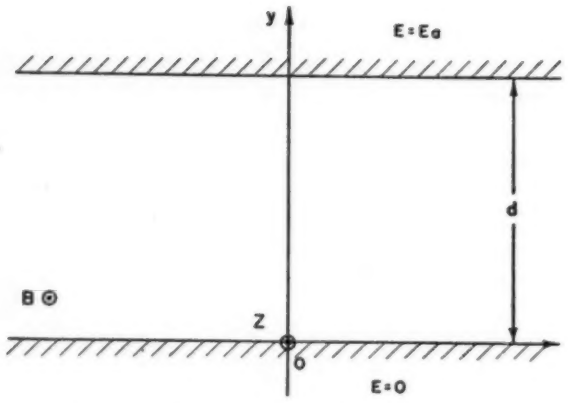


FIG. 1. Coordinate system for plane magnetron.

II. EQUATIONS FOR ELECTRON MOTION IN A MAGNETRON

Since we are going to apply the methods of statistical mechanics to the problem of electron motion in a magnetron, it is convenient to write the equations of motion in Hamiltonian form. The total energy of an electron is:

$$W=\left[\frac{1}{2m}\mathbf{p}-e\mathbf{A}\right]^2-eE, \tag{1}$$

where $\mathbf{p}$ is the momentum vector, $\mathbf{A}$ the vector potential of the constant magnetic field, and E the scalar electric potential. The momentum $\mathbf{p}$ is given by:

$$\mathbf{p}=m\mathbf{v}+e\mathbf{A}, \tag{2}$$

where $\mathbf{v}$ is the velocity vector of the electron.

Equation (1) defines the "energy state" of the electron with reference to an average potential produced by the electrode potentials and a "smeared out" space charge. In describing the motion of an electron in an arbitrary but fixed energy state we shall disregard entirely the fluctuations in space-charge density and electric field so that we can define an ideal phase space or μ-space for the electron motion. Then we can proceed to study the flow of the electrons between different energy levels in μ-space, caused by the interchange of energy between the electrons or, which is the same, by the fluctuations in space-charge density and potential.

To simplify the mathematical processes, we choose a plane magnetron and a Cartesian coordinate system (Fig. 1). The cathode is represented by the plane $y=0$, the anode by the plane $y=d$. The uniform magnetic flux vector $\mathbf{B}$ is parallel to the positive z-axis. The potential of the cathode is taken to be zero and the potential of the anode positive is equal to E_a. No variations are taking place with x and z,

$$\partial/\partial x=\partial/\partial z=0. \tag{3}$$

The vector potential $\mathbf{A}$ then has only one component, A_x, such that

$$-\partial A_x/\partial y=B_z=B, \tag{4}$$

$$A_x=-By, \tag{5}$$

since we can choose A as well as E to be zero at the cathode.

The Hamiltonian (Eq. (1)) then is

$$W=\frac{1}{2m}\{(p_x+eBy)^2+p_y^2+p_z^2\}-eE, \tag{6}$$

and the canonical equations of motion

$$-\frac{dp_x}{dt}=\frac{\partial W}{\partial x}=0=m\frac{dv_x}{dt}+eBv_y, \tag{7}$$

$$-\frac{dp_y}{dt}=\frac{\partial W}{\partial y}=\frac{1}{2m}(p_x+eBy)\cdot eB-e\frac{\partial E}{\partial y}=eBv_x-e\frac{\partial E}{\partial y}, \tag{8}$$

$$-\frac{dp_z}{dt}=\frac{\partial W}{\partial z}=0, \tag{9}$$

$$\frac{dx}{dt}=\frac{\partial W}{\partial p_x}=\frac{1}{m}(p_x+eBy)=v_x, \tag{10}$$

$$\frac{dy}{dt}=\frac{\partial W}{\partial p_y}=v_y, \tag{11}$$

$$\frac{dz}{dt}=\frac{\partial W}{\partial p_z}=v_z. \tag{12}$$

The *constants of the motion*, or the parameters that determine the energy state of each electron, are W, p_z and p_z. Optionally, we can specify the state of an electron by p_x, p_z and p_{y0}, the last quantity being the value of p_y at the cathode ($y=0$). It should be noted that p_{y0} may be imaginary, since the orbit of an electron may not necessarily reach the cathode.

In the study of the space charge in the magnetron we shall be interested in the distribution of the electrons in *phase space* or μ-space, i.e., a six-dimensional space with the coordinates x, y, z, p_x, p_y, p_z. Actually, a subspace y, p_x, p_y will contain all the boundary surfaces and orbit projections necessary for the study of the problem.

III. EQUILIBRIUM STATES

If the cathode of the magnetron is at constant temperature and no current flows to the anode, no energy is received or lost by the swarm of electrons in the tube. It should, under these hypothetic conditions, be possible to consider the magnetron as a closed system in thermal equilibrium. In this section we shall consider the conditions to be satisfied for thermal equilibrium in a magnetron.

On the other hand, when a minute current flows to the anode the problem becomes a transport problem, and

the solution is obtained from the thermal equilibrium by applying a small perturbation to the distribution function for the electrons. As the current is increased, the diffusion of electrons through the space charge from the cathode gradually changes to a steady flow of the whole space charge towards the anode. The intermediate conditions, with severely distorted distribution function but no coherent flow pattern, offer the greatest mathematical difficulties; unfortunately, there are reasons to believe that this is the actual state of affairs in a cut-off magnetron.

Since the space charge is assumed to be in equilibrium with the thermionic cathode emitter, the distribution density function at $y=0$ is necessarily of the Maxwell-Boltzmann type.

$$P_0=A\cdot\exp\left[-\frac{\alpha}{2m}(p_x^2+p_{y0}^2+p_z^2)\right]. \tag{13}$$

The distribution density function P at an arbitrary point approaches its equilibrium value under the influence of two simultaneous processes, convection by electron motion, and exchange of energy and momentum between the electrons because of Coulomb interaction.

We shall not include here the complete classical derivation of the distribution density function. It is only necessary to investigate the consequences of the magnetic field and the curvilinear orbits of the electrons.

Under equilibrium conditions the convection-current element formed by a certain group of electrons in phase space must be continuous throughout its path through x-y-z space. The expected value of a current element formed by electrons of y-directed momentum between p_{y0} and $p_{y0}+dp_{y0}$ at the cathode is, therefore, equal to the corresponding element at any other value of y,

$$eP_0p_{y0}dp_{y0}=ePp_ydp_y. \tag{14}$$

However, conservation of energy requires

$$W=\frac{1}{2m}\{p_x^2+p_{y0}^2+p_z^2\}$$
$$=\frac{1}{2m}\{(p_x+eBy)^2+p_y^2+p_z^2\}-eE, \tag{15}$$

and consequently,

$$p_{y0}dp_{y0}=p_ydp_y. \tag{16}$$

Continuity, therefore, requires that the distribution density at y is

$$Pdp_y=P_0dp_y. \tag{17}$$

It should be noted that the distribution density function P is given by Eq. (17) only for the regions in phase space accessible to electrons emitted from the cathode. We shall later map these regions.

The interaction between the electrons during thermal equilibrium must be such that the same number of electrons are removed from a certain cell in phase space as are entering into the cell during the same time interval.

Suppose that a certain encounter involves n electrons with the initial energy states $W_1\cdots W_n$, and after the encounter, the energy states $W_1'\cdots W_n'$. Conservation of energy requires that

$$W_1+W_2+\cdots W_n=W_1'+W_2'+\cdots W_n'. \tag{18}$$

The rates at which such encounters and the inverse encounters take place should be equal and are proportional to

$$P_1\cdot P_2\cdot P_3\cdots P_n=P_1'\cdot P_2'\cdots P_n'. \tag{19}$$

Since the exponential function transforms a sum into a product, it is obvious that the Maxwell-Boltzmann distribution function satisfies the requirements (18) and (19), as well as (13) and (17). Thus we can write:

$$P=A\cdot e^{-\alpha W}, \tag{20}$$

where A is a constant related to the total number of electrons in the space charge and $\alpha=1/kT$.

It is interesting to note that the presence of the magnetic field and the consequent curvature of the electron orbits do not affect the distribution function. The factors that determine the distribution arise from conservation of energy and isotropy of interaction.

In order to find a space-charge distribution compatible with thermal equilibrium, we integrate (20) with respect to p_x, p_y, and p_z from $-\infty$ or $+\infty$. The result is

$$\rho=\rho_0e^{\alpha eE}. \tag{21}$$

This relation is combined with Poisson's equation to

$$\frac{\partial^2E}{\partial y^2}=-\frac{\rho}{\epsilon_0}=-\frac{\rho_0e^{\alpha eE}}{\epsilon_0}, \tag{22}$$

$$\frac{1}{2}\left(\frac{\partial E}{\partial y}\right)^2=-\frac{\rho_0}{\alpha e\epsilon_0}e^{\alpha eE}+C_1. \tag{23}$$

In order to perform the second integration, we introduce

$$Q=e^{\alpha eE}, \tag{24}$$

$$dQ/Q(a+Q)^{\frac{1}{2}}=-2\rho_0\alpha e/\epsilon_0\,dy. \tag{25}$$

Since the first term in (23) is always positive, C_1 and a must be negative if a potential minimum exists.

Changing the sign of a and integrating, we get

$$2/(a)^{\frac{1}{2}}\tan^{-1}(Q-a/a)^{\frac{1}{2}}=by+C_2, \tag{26}$$

$$\rho=\rho_0\cdot a/\cos^2b(y-y_m)=\rho_0[\cos by_m/\cos b(y-y_m)]^2, \tag{27}$$

$$E=2/\alpha e\log\cos by_m/\cos b(y-y_m), \tag{28}$$

where

$$b=\left(\frac{-2\rho_0\alpha e}{\epsilon_0}\right)^{\frac{1}{2}}. \tag{29}$$

The boundary conditions are represented by the constants ρ_0, a, and y_m. For negative values of y_m the solution represents temperature-limited conditions.

Figure 2 indicates the space-charge distribution required for this thermal equilibrium. It obviously does not resemble the distribution in a cut-off magnetron, since it demands emission of electrons from the anode with the same temperature as those emitted from the cathode but with considerably higher density,

$$\rho_a = \rho_0 e^{\alpha e E_a}. \tag{30}$$

Nonetheless this equilibrium has a certain interest to us, since the difference between this distribution and the actual distribution in any particular volume element in real space may serve as a rough indication of the amount of diffusion that takes place in that element. A steady state is reached when the diffusion into every energy state is equal to the diffusion out of the same energy state.

Before we investigate more closely this diffusion process, we shall in the next section discuss the representation of electron energy states as points and orbits in phase space.

IV. INITIAL STATES AND PHASE-SPACE REPRESENTATION OF ELECTRON ENERGY STATES

When a magnetron is switched to a source with a voltage E smaller than the cut-off voltage of the magnetron, a charge Q will flow to the magnetron. The source supplies the energy EQ, but the stored energy in the magnetron is only $\frac{1}{2}EQ$. The difference is lost in the circuit resistance and, possibly, in a temporary increase in the electron temperature of the magnetron space charge above the cathode temperature. These transient effects we shall disregard and assume that the initial conditions are consistent with conservation of energy and momentum for each electron emitted from the cathode and with Poisson's law.

As stated earlier in this paper, the energy state of an electron is characterized by its energy W and two of the components of its momentum, p_x and p_z, these three

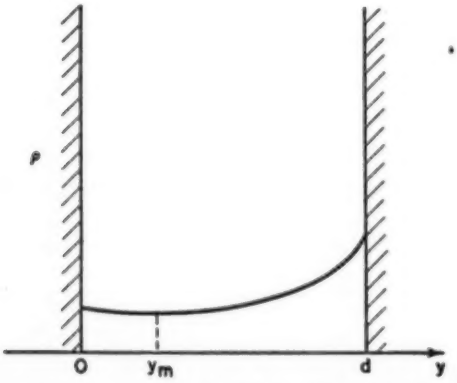


FIG. 2. Space-charge distribution at thermal equilibrium.

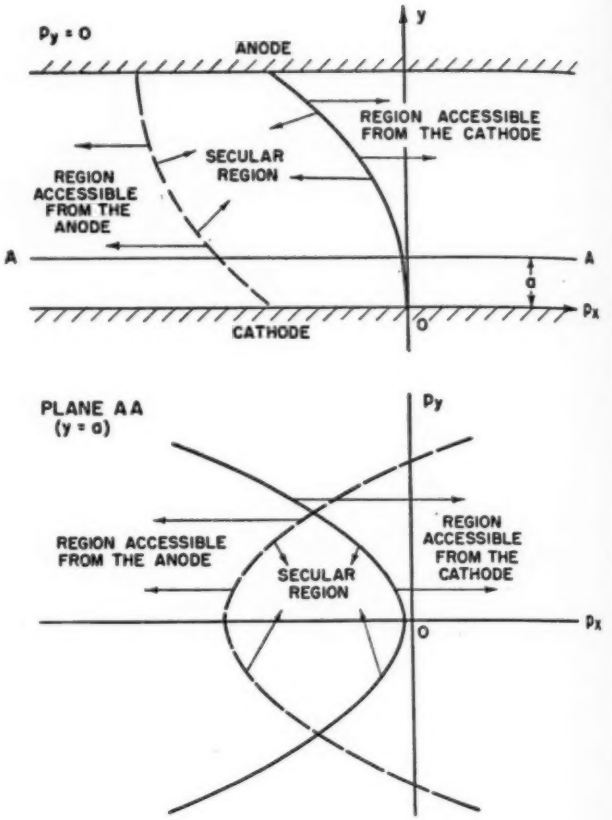


FIG. 3. Mapping of regions in phase space accessible to electrons whose orbits intersect the anode or cathode plane.

quantities being constants of the motion, as long as energy and momentum are conserved. On the other hand, the component p_y and the potential energy vary during the motion.

In order to map the regions in phase space accessible to the electrons emitted from the cathode it is sufficient to consider the space p_x, p_y, y.

The accessibility criterion is obtained from

$$2mW = p_x^2 + p_{y0}^2 + p_z^2 = (p_x + eBy)^2 + p_y^2 + p_z^2 - 2meE,$$

or

$$p_y^2 = p_{y0}^2 + 2meE - e^2B^2y^2 - 2p_x eBy, \tag{31}$$

where p_y and p_{y0} have to be real quantities.

Figure 3 shows a sketch of the boundary surface between the accessible and the inaccessible part of p_x-p_y-y-space. It is, of course, qualitative only, since $E(y)$ is not known but related to the space-charge distribution by Poisson's law. Integration of the distribution function (20) over the accessible space with respect to p_x and p_y would give the relation between ρ, E, and y that, combined with Poisson's equation, determines the initial conditions.

The inaccessible region has the shape of a distorted cone with its apex on the axis $p_y=0$, but in general not at $p_x=0$. The axis $p_y=0$ to the left of the apex is a generatrix of the surface. It should be pointed out that

the fact that the inaccessible region reaches the axis does not mean that electrons with $p_y=0$ and p_x in this range cannot escape from the cathode. They do, but only tangentially to the p_x-p_y plane, describe a broad elliptic orbit in the y-p_y plane and reach their turning point on the upper part of the boundary line of the accessible region or hit the anode.

In Fig. 3 the boundary for interception of the electrons by the anode is also indicated. In other words, any electron to the left of this surface will be removed by impact on the anode. The space-charge density can, therefore, be considered zero to the left of this boundary except where this region overlaps the region accessible from the cathode.

The volume between the two boundaries contains all the possible energy states whose electron orbits reach neither the anode nor the cathode. We shall refer to this volume as the *secular* region since the life of an electron energy state in this region is very long compared with the period of the cyclic motion of the electron. It should be noted that:

1. *Only an infinitesimal change in momentum is required for an electron to cross the boundary into this volume.*

2. *The electron population of this volume will continue to increase until the current flowing to the anode through the opposite boundary equals the current entering the volume from the cathode-accessible region.*

3. *Whether the discrete electron-electron interaction is weak or strong determines primarily the time required to reach a steady state but not necessarily the final space-charge distribution.*

Before discussing the possibility of determining at least roughly the final space-charge distribution, we shall give some brief comments on previously proposed solutions to the space-charge distribution in the cut-off magnetron. Common to them all is that the solution is identified with some form of what we have called initial conditions; sometimes the justification is given that the interaction between discrete electrons is small enough to be neglected. The three underlined conditions above show that this reasoning is not correct because of the effect of the interaction, however weak, is cumulative. In other words, these distributions are unstable, since the perturbations due to electron interaction do not produce fluctuations about the initial state, but a drift away from this state.

The Brillouin or single-stream space-charge distribution[1–3] in phase space (Fig. 4) is limited to the line AB and independent of y within this range. At first sight this distribution may appear to be a possible equilibrium at 0°K temperature. However, it would be strange

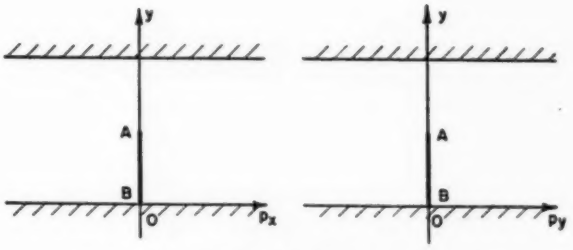


FIG. 4. Brillouin distribution in phase space.

indeed if all electrons occupied the same point in momentum space, although lower energy states certainly are possible. But we must consider the possibility that the state be metastable because of the lack of a process whereby the unoccupied energy states could be populated. Obviously such a process does exist. Since the space charge is formed by discrete electrons whose velocity varies with y, even though their orbits do not intersect, the electric field at the edge of the space-charge cloud necessarily fluctuates in magnitude and direction both in time and space. A diffusion will therefore take place, both out into the unoccupied space and towards the cathode. The energy required for this random motion is, of course, supplied by the dc electric field.

The double-stream distribution described by Slater,[4] Page and Adams,[5,6] and others is represented by an approximately elliptic line charge in phase space (Fig. 5). The space-charge density in real space at the cathode and at the edge of the swarm is infinite; a finite minimum is located at an intermediate plane. This is also a 0°K distribution, since all electrons have the same energy although occupying a line rather than a point in momentum space. There can be no question about a metastable state in this case, since the electron orbits intersect with considerable relative velocity so that exchange of energy and momentum is inevitable.

Twiss[7] has considered the modification in the initial space-charge distribution produced by the initial velocities of the electrons. The result is essentially the initial state mentioned earlier in this section as obtain-

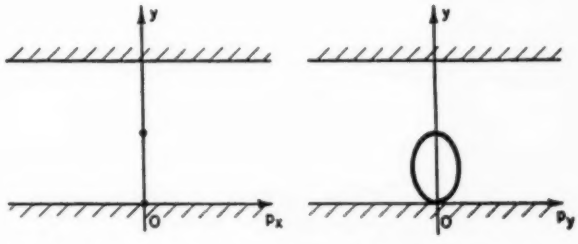


FIG. 5. Double-stream distribution in phase space.

[1] L. Brillouin, "Electronic theory of the plane magnetron," *Advances in Electronics* (Academic Press, Inc., New York, 1951), Vol. 3, p. 85.

[2] L. Brillouin and F. Bloch, "Electronic theory of the cylindrical magnetron," *Advances in Electronics* (Academic Press, Inc., New York, 1951), Vol. 3, p. 145.

[3] A. W. Hull, Phys. Rev. **18**, 31 (1921).

[4] J. C. Slater, *Microwave Electronics* (D. Van Nostrand Company, Inc., New York, 1950).

[5] L. Page and N. J. Adams, Phys. Rev. **69**, 492 (1946).

[6] L. Page and N. J. Adams, Phys. Rev. **69**, 494 (1946).

[7] R. Q. Twiss, *On the Steady-State and Noise Properties of Linear and Cylindrical Magnetrons*, doctoral thesis (Massachusetts Institute of Technology, 1949).

able by integration of the distribution function (20) over the accessible region of momentum space. It avoids the discontinuities and singularities of the two distributions mentioned above, but is otherwise subject to the same criticism; it is an initial state but not a steady state.

V. DISCUSSION OF THE MAGNETRON SPACE-CHARGE DISTRIBUTION AS A DIFFUSION PROBLEM

A calculation of the steady state in the cut-off magnetron from the initial state previously indicated is complicated by the following circumstances:

1. The relationships governing the transfer of momentum between the electrons, including close encounters as well as distant encounters, are rather involved even when the state of the space charge differs very little from thermal equilibrium.

2. The actual distribution density is very irregular because of the initially abrupt variation at the accessibility boundaries.

3. The distribution density function is furthermore distorted by the automatic "sorting" of electrons that have gained or lost energy, respectively, in the process of interaction. Most of the electrons that have gained energy will return to the cathode, while those that have lost energy will not. On the average, therefore, the returning electrons will have a slightly higher temperature than those leaving the cathode. (We use the term "temperature" for convenience although the distribution is not regular Maxwell-Boltzmann.)

We shall here try to arrive at a qualitative understanding of the way in which a steady state is reached without resolving quantitatively the complications enumerated above.

The drift of electrons through phase space is such that if the system were temporarily closed its state would gradually approach thermal equilibrium.

During thermal equilibrium the number of electrons that move from one particular volume element in phase space to another such element is equal to the number that moves in the opposite direction. We should expect it to be possible to calculate the instantaneous value of the interaction current from one volume element in the phase space to another by comparing the instantaneous distribution with the equilibrium distribution. Here, however, we should not consider thermal equilibrium, which is determined by the boundary conditions of the whole space charge, but the standard normal distribution of (p_x+eBy), p_y and p_z that has the same total population, the same energy, and the same center of gravity as the initial distribution. The total interaction current across a certain boundary surface in momentum space would be obtained by a double integration over these two distribution density functions. In addition to a function of the densities of the two volume elements in both distributions the integrand must necessarily contain a weighting function expressing the probability of the required change of momentum between the two volume elements. At least two simultaneous asymptotic processes should be considered that give quite different weighting functions. The first one operates by close binary encounters and is a discrete random process of well-known type. The influx to a certain volume element in phase space because of its density deficiency is largely determined by the distance of the element from the center of gravity of the distribution density functions. In this case the weighting function accounts for the impact parameter and the angle between the relative velocities of the two electrons with respect to their common center of gravity.

The second process is the result of interaction between a large number of electrons at considerably larger distances. The weighting function in this case permits only very small changes of momentum. The interaction current density across a certain surface in momentum space is consequently determined chiefly by a relation between the actual density gradient at this surface and the equilibrium density gradient there.

It has been shown[8] that when the interaction is due to Coulomb forces, neither one of these two processes can in general be neglected in comparison with the other.

In enumerating the factors that determine the steady state we should begin with the boundary conditions at the cathode. It is natural to assume that the mass and thermal capacity of the cathode are so large that the temperature and distribution of the emitted electrons are independent of the temperature and distribution of the returning electrons, that is to say, the energy and momentum of the electron cloud are not conserved at the cathode. If momentum were at least partly conserved, the center of gravity of the distribution could not fall on the line $p_x=-eBy$ but somewhere between this line and $p_x\doteq 0$. The assumption made here appears satisfactory as long as the emission is strictly thermionic; if secondary emission is appreciable, the boundary conditions become much more difficult to state.

Between the cathode and the potential minimum the inaccessible part of the phase space is likely to be small and not to include dense regions of the distribution. The potential and space-charge density then drops approximately according to (21) until roughly the Brillouin density is reached.

The most interesting region in real space is between the potential minimum and the edge of the cloud. The boundary of the secular volume in phase space on the cathode side is likely to be almost parallel to the y-axis here, since the potential distribution is probably not radically different from that of the Brillouin solution (see Fig. 6). The space-charge density in the secular volume obviously must be such that the net number of electrons entering from the right equals the net number leaving at the left boundary. This density is obviously smaller than the one required for ideal thermal equilibrium; the center of gravity of the distribution must,

[8] Cohen, Spitzer, and Routly, Phys. Rev. **80**, 230 (1950).

therefore, be expected to be to the right of the line $p_x=-eBy$, that is, closer to the right-hand boundary than to the left-hand one. The space-charge density required for a steady state is therefore closer to the density at the right boundary than to the one at the left, which is close to zero. To calculate this space-charge distribution is evidently very difficult, but it should be clear that its density is by no means negligible as far as calculation of space-charge distribution in real space is concerned.

Twiss maintains that a double-stream motion with considerably more than thermal energy exists here in the cathode-accessible volume of phase space. The square of the y-directed momentum of an electron can be written (31)

$$p_y^2=p_{y0}^2+2mE-e^2B^2y^2-2p_xeBy\cong p_{y0}^2-2p_xeBy. \quad (32)$$

The orbits of the electrons in the p_y-y-plane are parabolas, and since $eBy\gg p_x$, p_y may be of a different order of magnitude than p_{y0} for negative values of p_x. The orbits do not extend to the anode, because outside the edge of the space-charge swarm $e^2B^2y^2$ is considerably larger than $2mE$, so that the electrons turn back into the cloud.

At first sight these conclusions seem to be inconsistent with the view presented in Sec. IV, that the initial energy distribution at any point is an incomplete Maxwell-Boltzmann distribution of cathode temperature. The answer is that as the electrons move in the positive y-direction and gain kinetic energy from the electric field, they also move towards the high energy fringes of a distribution that, if it were complete, would have a much higher space-charge density than at the cathode.

The high energy electrons will be much closer to the center of gravity of the actual, considerably distorted distribution than to the center of gravity of the Maxwell-Boltzmann distribution of which they form a small part. Consequently, the average kinetic energy or "temperature" of the actual distribution will be much higher than at the cathode.

This increase of the average electron energy with y no doubt contains the clue to a number of magnetron problems. Considering the cut-off magnetron as a resistor, we should expect the increased electron temperature to increase the noise output. Twiss has attributed the noise level to a "noise amplification" caused by the double-stream interaction in the space-charge cloud. The points of view may not be equivalent, but they are certainly related.

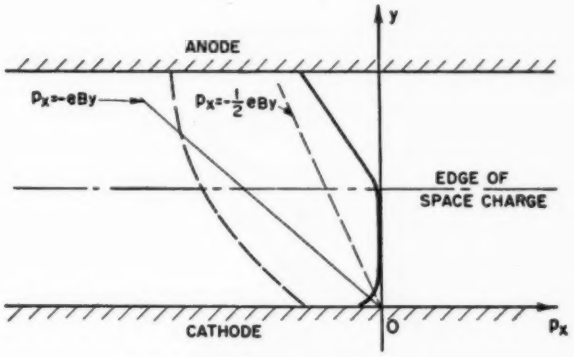


FIG. 6. Space-charge-limited conditions.

The difference between the space-charge distribution discussed here and the Brillouin or Twiss solution is probably not very large when the volume of the secular region in phase space is small, i.e., when the tube is operated close to the cut-off voltage. The anode current will be appreciably larger than the direct convection current from cathode to anode, however, because of the diffusion through the secular region.

When the anode voltage is far below the cut-off voltage, on the other hand, the secular volume is large, and the total space charge there may be large enough to affect the space-charge density in real space and the potential distribution appreciably.

The mathematical difficulties discourage any attempt to predict numerically the space-charge distribution in a cut-off magnetron; we must, therefore, look to experimental investigations for quantitative information. Reverdin[9] has described an interesting electron-optical method to explore the magnetron space charge. His results show that the steady-state formation is sensitive to cathode geometry and temperature, but his data are too scanty to yield any definite quantitative information as to the effect of the various parameters of the problem.

[9] D. L. Reverdin, J. Appl. Phys. **22**, 257 (1951).

The Determination of Complex Dielectric Constants of Absorptive Liquids by Microwave Interferometry*†

FRANKLIN H. BRANIN, JR.‡
Frick Chemical Laboratory, Princeton University, Princeton, New Jersey
(Received April 7, 1952)

This paper describes both a traveling-wave method and a standing-wave method for determining complex dielectric constants of liquids from measurements made at 10-cm wavelength with a traveling probe immersed in a dielectric-filled slotted line. An interferometric null technique was developed for improving the accuracy of the measurements, especially in highly absorptive media. In the traveling-wave method, absolute measurements of both wavelength and absorption index can be made simultaneously; from these two quantities the complex dielectric constant is readily computed. In the standing-wave method, the absorption index is determined graphically from the power-standing-wave ratio while the wavelength is obtained by a modified null technique. For low loss media, a simplified standing-wave method is practical whereby the absorption index is obtained from measurements of the breadths of successive minima of the standing wave. Corrections for the finite resistance of the short circuit used in the standing-wave method and for conductor losses along the slotted line are fully treated.

INTRODUCTION

FROM the standpoint of the molecular theory of dielectrics, the many recently developed microwaves techniques offer a useful tool for the investigation of molecular behavior in a region of the frequency spectrum which has been but little explored heretofore.§ The present research was undertaken as part of a systematic study of dielectric dispersion and absorption at microwave frequencies in a number of polar organic liquids.†[1–8] Results of the measurements made by the method described in this paper are reported elsewhere.[9]

In the determination of complex dielectric constants by microwave propagation techniques, it is only the propagation characteristics of the wave and not the dielectric properties of the medium which can be observed. Therefore, the problem reduces to one of making measurements which will lead as directly as possible to the evaluation of the complex propagation constant, $\gamma=\alpha+j\beta$, of the wave within the medium. Once this has been accomplished, the computation of the complex dielectric constant, $\epsilon^*=\epsilon'-j\epsilon''$, is a simple matter. In the present work, the necessary measurements were made directly on either a traveling wave or a standing wave within the medium by means of a probe traveling along a slotted section of coaxial transmission line filled with the dielectric under investigation. Although the procedures were worked out for coaxial line only at a wavelength of 10 cm, the principle of the method is applicable to wave-guide measurements.

Of the numerous methods already known for the microwave determination of complex dielectric constants,[1,2,6,10–17] only a few have involved measurements made directly within the medium.[10,12,17] The advantage of such direct methods is the relative simplicity of the equations required for the computation of ϵ^* from the observed data. The standing wave method first described by Drude,[10] and more recently by Brown, Hoyler, and Bierwerth,[12] is the basis of one of the present methods where an improvement has been made in the wavelength measurement by the use of an interferometric null technique. This technique greatly increases the accuracy of the measurement, particularly in highly absorptive media. Moreover, when applied to the traveling-wave method, the interferometric technique permits the absolute measurement of both the wavelength and the absorption index to be made simultaneously.

* This research was supported by the ONR. Reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government is permitted.

† This paper is based upon a dissertation presented in January, 1950, to the Faculty of Princeton University in candidacy for the Degree of Doctor of Philosophy.

‡ Present address: Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico.

§ For the most part these techniques stem from earlier damped-wave methods proposed fifty or more years ago. Recent improvements consist largely in substituting continuous waves for damped waves and modern receivers for the crude detectors of that day.

[1] Heston, Hennelly, and Smyth, J. Am. Chem. Soc. **70**, 4093 (1948).
[2] H. L. Laquer and C. P. Smyth, J. Am. Chem. Soc. **70**, 4097 (1948).
[3] Hennelly, Heston, and Smyth, J. Am. Chem. Soc. **70**, 4102 (1948).
[4] W. M. Heston, Jr., and C. P. Smyth, J. Am. Chem. Soc. **72**, 99 (1950).
[5] Heston, Hennelly, and Smyth, J. Am. Chem. Soc. **72**, 2071 (1950).
[6] Heston, Franklin, Hennelly, and Smyth, J. Am. Chem. Soc. **72**, 3443 (1950).
[7] Franklin, Heston, Hennelly, and Smyth, J. Am. Chem. Soc. **72**, 3447 (1950).
[8] Curtis, McGeer, Rathmann, and Smyth, J. Am. Chem. Soc. **74**, 644 (1952).
[9] F. H. Branin, Jr. and C. P. Smyth, J. Chem. Phys. Soc. **20**, 1121 (1952).
[10] P. Drude, Z. physik. Chem. **23**, 267 (1897).
[11] S. Roberts and A. von Hippel, J. Appl. Phys. **17**, 610 (1946).
[12] Brown, Hoyler, and Bierwerth, *Radio Frequency Heating* (D. Van Nostrand Company, Inc., New York, 1947), p. 218.
[13] C. G. Montgomery, *Techniques of Microwave Measurements* (McGraw-Hill Book Company, Inc., New York, 1948), Chap. 10.
[14] T. W. Dakin and C. N. Works, J. Appl. Phys. **18**, 789 (1947).
[15] W. H. Surber, Jr., J. Appl. Phys. **19**, 514 (1948).
[16] W. H. Surber, Jr., and G. E. Crouch, Jr., J. Appl. Phys. **19**, 1130 (1948).
[17] Collie, Hasted, and Ritson, Proc. Phys. Soc. (London) **60**, 145 (1948).

The relation between the complex dielectric constant of an absorptive dielectric and the complex propagation constant of a wave being propagated within the medium is derived from the appropriate Maxwell equations,

$$\nabla\times\mathbf{E}=-j\omega\mu_0 H, \tag{1}$$

$$\nabla\times\mathbf{H}=(\omega\epsilon''\epsilon_0+j\omega\epsilon'\epsilon_0)\mathbf{E}, \tag{2}$$

where $\epsilon_0=(1/36\pi)\times10^{-11}$ farad cm^{-1} and $\mu_0=4\pi\times10^{-9}$ henry cm^{-1}; $\mathbf{E}$ and $\mathbf{H}$ are the instantaneous electric and magnetic field vectors. The simplest solution of these equations, corresponding to plane-wave propagation along the coordinate x, is

$$\mathbf{E}=\mathbf{E}_0e^{-\gamma x+j\omega t}, \tag{3}$$

$$\mathbf{H}=\mathbf{H}_0e^{-\gamma x+j\omega t}. \tag{4}$$

From these equations and the definition, $\gamma=\alpha+j\beta$, where α is the attenuation constant and β is the phase constant, it is readily shown that

$$\gamma^2=(\alpha^2-\beta^2)+2j\alpha\beta-\omega^2\epsilon'\epsilon_0\mu_0+j\omega^2\epsilon''\epsilon_0\mu_0. \tag{5}$$

Next, using the relation $c=(\epsilon_0\mu_0)^{-\frac{1}{2}}$ where $c=\omega\lambda_0/2\pi$ is the velocity of the wave in free space and λ_0 is the free space wavelength, it follows from Eq. (5) that

$$\epsilon'=(\lambda_0/\lambda_d)^2[1-(\alpha/\beta)^2], \tag{6}$$

$$\epsilon''=(\lambda_0/\lambda_d)^2[2\alpha/\beta], \tag{7}$$

where $\lambda_d=2\pi/\beta$ is the wavelength within the dielectric.

Equations (6) and (7) apply not only to plane waves but also to the principal or TEM mode of propagation in a coaxial transmission line and are fundamental for the present discussion. For wave-guide propagation and for higher modes of propagation in a coaxial line, however, although Eq. (7) is directly applicable, Eq. (6) must be modified as follows:

$$\epsilon'=(\lambda_0/\lambda_d)^2[1-(\alpha/\beta)^2]+(\lambda_0/\lambda_c)^2, \tag{8}$$

where λ_c is the cut-off wavelength of the particular mode of propagation in vacuum. Since Eqs. (6) and (7) are applicable to plane waves, they are obviously the microwave counterparts of the familiar optical relations, $\epsilon'=n^2(1-\kappa^2)$ and $\epsilon''=n^2(2\kappa)$, or in complex form,

$$\epsilon^*=(n^*)^2=[n(1-j\kappa)]^2, \tag{9}$$

where $n=\lambda_0/\lambda_d$ is the refractive index and $\kappa=\alpha/\beta$ is the absorption index of the medium.

EXPERIMENTAL METHODS

In the direct measurement of microwave propagation constants by means of a probe traveling along a dielectric-filled slotted line, it can be assumed that the probe signal is proportional to and differs by a constant amount in phase from the instantaneous electric field at each probe position, provided that the probe intercepts a constant fraction of the electric field energy and exerts a negligible perturbation on the wave within the slotted line. Thus, if only an incident wave is present, the probe signal is characterized by Eq. (3), whereas if both an incident and a reflected wave are present, the probe signal is proportional to the vector sum of the two waves at each point. These statements apply to uniform wave guides and transmission lines of any configuration provided that the entire electric field of such configuration is immersed in the dielectric. If an appreciable fraction of the total electric field energy is exterior to the medium, errors will be introduced. For this reason, the slot which permits entry of the probe into the dielectric must be both narrow and deep and must be filled to a reasonable depth with dielectric.

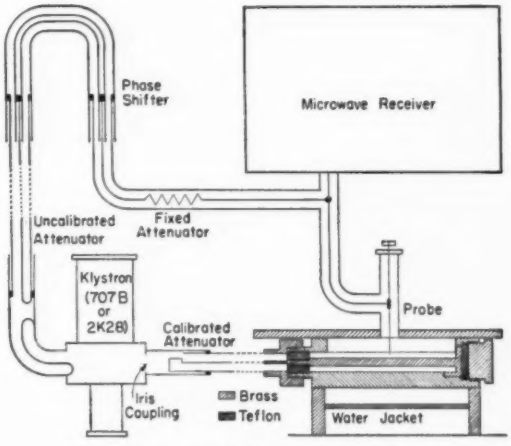


FIG. 1. Functional diagram of apparatus.

Traveling-Wave Method

If the dielectric-filled slotted line is terminated in a matched load so that sensibly all of the incident wave energy is absorbed, the probe signal obeys Eq. (3); and hence both α and β may be determined directly from the observed amplitude and phase changes of the probe signal. The measurement of amplitude changes is simple enough, but the measurement of phase requires the use of a reference or comparison signal of adjustable amplitude and phase. The apparatus used for these measurements is depicted schematically in Fig. 1.‖ Essentially, this apparatus is a microwave interferometer because it depends upon the interference of the probe signal and the reference signal to produce a null relative to which the phase and amplitude changes of the probe signal can be measured.

The signal level in the slotted line, and hence in the probe, is controlled by a calibrated piston attenuator[18–22]

‖ The short-circuit termination shown for the slotted line is for use with the standing-wave method. In the traveling-wave method, a matched termination must be used instead.

[18] D. E. Harnett and N. P. Case, Proc. Inst. Radio Engrs. **23**, 57 (1935).

[19] R. E. Grantham and J. J. Freeman, Am. Inst. Elec. Engrs., Tech. Paper 4–50 (1947); Elec. Eng. **67**, 535 (1948).

[20] J. J. Freeman, J. Research Natl. Bur. Standards **40**, 235 (1948).

[21] H. A. Wheeler, Monograph No. 8, Wheeler Laboratories, Inc., Great Neck, New York.

[22] See reference 13, Chap. 11.

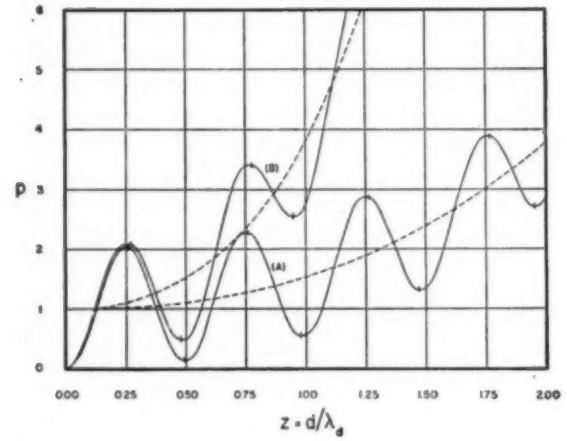


FIG. 2. Power-standing-wave pattern in short-circuited line filled with an absorptive dielectric. (A) $\kappa=0.08$; (B) $\kappa=0.16$. (Dashed line is $\cosh 4\pi\kappa z$.)

which is iris-coupled into the cavity of the klystron oscillator. The reference signal is loop-coupled into the oscillator cavity and is adjusted in amplitude by an uncalibrated piston attenuator and in phase by a "line stretcher." Thus, for any position of the probe, a null can be produced by proper manipulation of these amplitude and phase adjustments. Flexible coaxial cables conduct the probe signal and reference signal to a T-junction at the input to a microwave receiver which responds to the vector sum of the two signals at the T-junction. Since the receiver is used solely as a null detector in this measurement, it may be operated at full sensitivity in order to permit maximum precision in making the null adjustments.

The procedure for measuring the wavelength and absorption index is as follows: First, a null is obtained by properly adjusting the reference signal controls with the probe placed at any convenient position near the input end of the slotted line. Next, the probe is moved down the line to a new position where the null can be restored simply by adjusting the signal level in the slotted line using the calibrated piston attenuator.¶ Since the piston attenuator produces attenuation without change of phase, it follows that the probe signal must change phase by 360° between null points; accordingly, the probe displacement must be exactly λ_d. At the same time, because the attenuation produced by a probe displacement of one wavelength is $e^{-\alpha\lambda_d}$ or $e^{-2\pi\kappa}$, a corresponding change in signal level must be introduced by the piston attenuator in restoring the second null. Moreover, since the piston attenuator is a wave guide operating beyond cutoff, it produces an exponential change of amplitude with a linear displacement of the pick-up loop. Therefore, the pick-up loop displacement y required to restore the second null is directly proportional to the absorption index of the dielectric, the appropriate equation being

$$\kappa=y(1/\lambda^2_c-1/\lambda^2_0)^{\frac{1}{2}}=0.01831A, \qquad (10)$$

where λ_c is the cut-off wavelength of the attenuator in vacuum and A is the attenuation in decibels. In this procedure the measurements of both λ_d and κ are absolute because the first measurement depends solely upon a linear displacement of the probe, while the second depends solely upon a linear displacement of the attenuator pick-up loop.

The major difficulty encountered in applying the traveling-wave method to the measurements of interest in this investigation was that of providing a matched termination suitable for use with a wide variety of dielectrics. In testing this method experimentally, a double stub tuner and a resistive load, connected to a specially designed output fitting for the slotted line, were used to absorb the incident wave energy, but this scheme proved impractical for general use. Because of limitations on the volume of material available, it was not possible to use a tapered resistive termination,[23] nor was a quarter-wave step matched load practical[23] because of wavelength variations from medium to medium. In highly absorptive dielectrics, however, such as water and nitrobenzene, the reflected wave is so rapidly attenuated that the traveling-wave method becomes feasible with almost any type of termination. Nevertheless, for greater convenience in making the desired measurements, it proved better to terminate the slotted line with a short circuit and to use a modification of the null technique in conjunction with the standing-wave method.

TABLE I. Power-standing-wave ratio as a function of absorption index.

κ	PSWR (db)	$\log\kappa+\frac{1}{2}\log$PSWR	$\frac{1}{4}-z_{\min}$
0.00	00	9.50285–10	0.000000
0.01	30.0570	9.50285–10	0.000050
0.02	24.0369	9.50287–10	0.000200
0.03	20.5153	9.50289–10	0.000452
0.04	18.0172	9.50292–10	0.000807
0.05	16.0799	9.50296–10	0.001268
0.06	14.4974	9.50302–10	0.001836
0.07	13.1600	9.50310–10	0.002517
0.08	12.0022	9.50320–10	0.003313
0.09	10.9819	9.50334–10	0.004231
0.10	10.0699	9.50350–10	0.005276
0.11	9.2462	9.50370–10	0.006453
0.12	8.4954	9.50395–10	0.007770
0.13	7.8064	9.50426–10	0.009232
0.14	7.1698	9.50462–10	0.010849
0.15	6.5797	9.50508–10	0.012628
0.16	6.0295	9.50560–10	0.014577
0.17	5.5155	9.50623–10	0.016707
0.18	5.0336	9.50695–10	0.019026
0.19	4.5790	9.50770–10	0.021544
0.20	4.1552	9.50879–10	0.024272
0.21	3.7542	9.50993–10	0.027223
0.25	2.3680	9.51634–10	0.041502
0.30	1.0616	9.53020–10	0.066311

¶ Constant sensitivity in locating the null points is maintained by leaving the reference signal level constant.

[23] See reference 13, Chap. 12.

Standing-Wave Method

When the dielectric-filled slotted line is terminated by a short circuit, substantially all of the incident wave energy is reflected and a standing wave is set up. Let us designate the electric intensities of the incident and reflected waves at a distance, d, from the short circuit by the expressions $\mathbf{E}^i{}_d = \mathbf{E}^i{}_0 e^{+\gamma d}$ and $\mathbf{E}^r{}_d = \mathbf{E}^r{}_0 e^{-\gamma d}$, where $\mathbf{E}^i{}_0$ and $\mathbf{E}^r{}_0$ are the intensities at the short circuit. If we now assume that the resistance of the short circuit is zero, then it follows that $\mathbf{E}^i{}_0 = -\mathbf{E}^r{}_0$ and that the standing wave, which is the vector sum of the incident and reflected waves, is given by the expression,

$$\mathbf{E}_d = \mathbf{E}^i{}_d + \mathbf{E}^r{}_d = \mathbf{E}^i{}_0(e^{+\gamma d} - e^{-\gamma d}) = 2\mathbf{E}^i{}_0 \sinh\gamma d. \quad (11)$$

The standing wave may be more conveniently described in terms of the relative power, defined as $p = |\mathbf{E}_d/2\mathbf{E}^i{}_0|^2$, since the probe signal is considered in terms of relative power. Equation (11) is readily transformed into the expression

$$p = \cosh 2\alpha d - \cos 2\beta d = \cosh 4\pi\kappa z - \cos 4\pi z, \quad (12)$$

where $z = d/\lambda_d$ is the distance from the short circuit in fractional wavelengths.

Equation (12) is plotted in Fig. 2 for two values of κ; it will be observed that as the absorption index increases, the standing-wave minima become less pronounced and eventually vanish. When $\kappa > 0.33$, even the first minimum vanishes and the standing-wave method is no longer applicable. In such a case, the traveling-wave method must be used.

The determination of the absorption index from the standing-wave pattern depends upon the measurement of the power-standing-wave ratio, PSWR, defined as the ratio of the power at the first maximum to that at the first minimum from the short circuit. The relationship between κ and PSWR has been calculated by successive approximations from Eq. (12) using 10-place tables,[24] and the results are reported in Table I; fortunately, a graph of $\log\kappa$ *versus* PSWR in decibels is almost a straight line as indicated by the slowly changing value of the quantity $\log\kappa + \frac{1}{2}\log$ PSWR tabulated in column 3. This fact permits graphical interpolation to be made with an accuracy commensurate with PSWR measurements of >0.01-db accuracy. In column 4, the distance between the first minimum and the half-wave point is tabulated as a function of κ.

PSWR is measured by the substitution method[25] using the calibrated piston attenuator of Fig. 1. In this measurement, the reference signal must be fully attenuated. Thus, the microwave receiver no longer acts as a null detector and its primary requirement is stability rather than sensitivity. Linearity of the receiver is not necessary, however, since the substitution method always insures that a constant signal level

[24] Table of Hyperbolic Sines and Cosines, Federal Works Administration, New York (1941).

[25] See reference 13, pp. 570, 915.

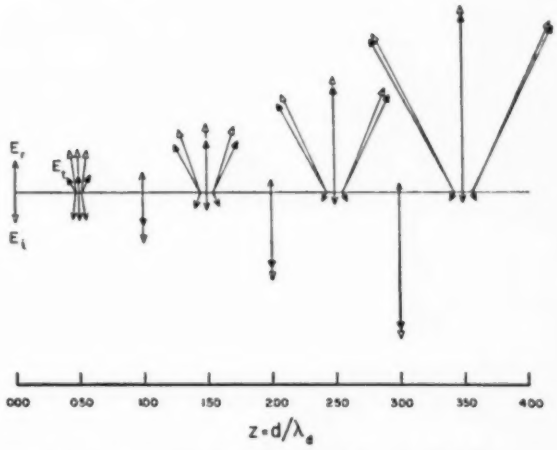


FIG. 3. Vector representation of incident and reflected waves in short-circuited line filled with an absorptive dielectric. ($\kappa = 0.08$.) $\mathbf{E}_i$ = Incident electric wave; $\mathbf{E}_r$ = reflected electric wave; $\mathbf{E}_t$ = electric standing wave.

appears at the receiver input, whether the probe is at the maximum or at the minimum. PSWR is read off the attenuator dial which is calibrated in decibels.**

The basis for the interferometric measurement of wavelength is a one-to-one correspondence between a 360° phase shift in the probe signal and a probe displacement of exactly λ_d. Although this correspondence holds at all points in a traveling wave, only certain sets of points satisfy this requirement in a standing wave. From Fig. 3, which shows the incident and reflected electric waves in terms of vectors, it is clear that the foregoing correspondence condition is fulfilled when the incident and reflected waves are in phase opposition.†† Therefore, if these points—specifically the set of points $z = N + \frac{1}{2}$, where $N = 0, 1, 2, 3$, etc.—can be located, the measurement of λ_d by the interferometric null technique is feasible. These points may be located by successive approximations according to the following procedure.

The position of the first minimum is established and the location of the point $z = 0.5$ is estimated by applying the appropriate correction factor from column 4, Table I. A null is then produced by adjusting the reference signal, and the probe is moved into the vicinity of the point $z = 1.5$ to a position where the null may be restored simply by adjusting the amplitude of the reference signal.‡‡ The probe is next moved to a position in the vicinity of $z = 2.5$ where the null may be similarly restored; the process may be repeated at

** A small frequency correction is required (see reference †) as indicated by Eq. (10).

†† The condition is also satisfied when the incident and reflected waves are in phase with each other; but at these points, which are near the standing-wave maxima, the rate of change of phase of the standing wave is so small that the position of an interference null cannot be located with adequate precision.

‡‡ In the standing-wave method, the amplitude of the reference signal is varied—i.e., *increased* at successive null points—in order to compensate for the decreasing rate of phase change of the probe signal. This helps to maintain a fairly constant sensitivity in locating the null points.

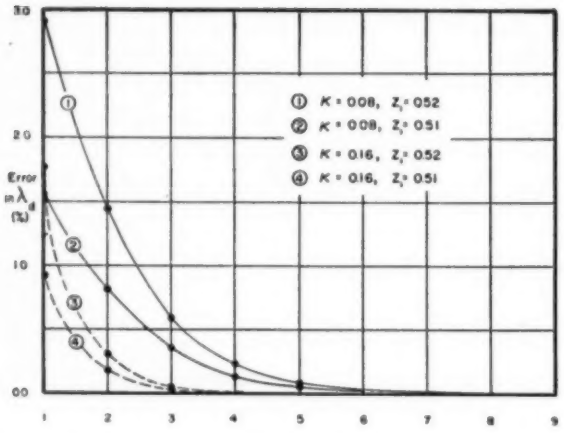


FIG. 4. Convergence of successive estimates of λ_d in the standing-wave method. (z_1 is the estimated position of the first half-wavelength point. Curves for $z_1<0.50$ are similar to those shown.)

succeeding odd half-wavelength points, if desired. Finally, the successive probe displacements are intercompared.

If the null points are equidistant, then they correspond exactly to the set $z=N+\frac{1}{2}$, and the probe displacements are all exactly one wavelength, λ_d. This case is illustrated by the colinear vectors in Fig. 3. Usually, however, the position of the first half-wavelength point is either underestimated or overestimated with the result that successive probe displacements between points of the standing wave differing in phase by 360° are either less than or greater than λ_d, as illustrated in Fig. 3. Fortunately, a rapid convergence toward the actual value of λ_d is obtained as shown in Fig. 4 so that in most cases, after suitable correction of the first estimated position of $z=0.5$, a second set of probe displacements yields the wavelength within the desired precision. With practice a third trial seldom becomes necessary.

Although this procedure of successive approximations is somewhat cumbersome, the gain in accuracy and precision over the measurement of wavelength on the basis of the distance between minima of the standing wave[10,12] more than compensates for the added complication.

Short-Circuit Corrections

In the foregoing discussion, it has been expressly assumed that the resistance of the short-circuit termination for the slotted line is zero. Actually, the resistance is finite and significant corrections are required in the determination of the absorption index, particularly for large values of PSWR; no correction is needed in the wavelength measurement, however.

If Z_0 is the characteristic impedance of the slotted line in vacuum—or in air, the difference being negligible—then its characteristic impedance Z^*, when it is filled with an absorptive dielectric, becomes

$$Z^*=Z_0/\epsilon^{*\frac{1}{2}}=Z_0/[n(1-j\kappa)]. \tag{13}$$

Assuming that the short-circuit impedance is purely resistive, namely $Z_s=R_s$, the complex reflection coefficient $\Gamma^*=\Gamma'-j\Gamma''$ of the short circuit is

$$\Gamma^*=\frac{R_s-Z^*}{R_s+Z^*}=\frac{(R_s\epsilon^{*\frac{1}{2}}/Z_0)-1}{(R_s\epsilon^{*\frac{1}{2}}/Z_0)+1}. \tag{14}$$

It is convenient to define the small quantity $\Delta^*=\Gamma^*+1=\Delta'-j\Delta''$ and to express it in terms of the normalized short-circuit impedance, $r^*=r'-jr''=nR_s/Z_0-jn\kappa R_s/Z_0$. Thus, since $r''=\kappa r'$ and $r'\ll 1$, it may be shown that

$$\Delta'=\frac{2r'[1+r'(1+\kappa^2)]}{1+2r'[1+\frac{1}{2}r'(1+\kappa^2)]}\approx 2r' \tag{15}$$

$$\Delta''=\frac{2\kappa r'}{1+2r'[1+\frac{1}{2}r'(1+\kappa^2)]}\approx 2\kappa r'. \tag{16}$$

Next, using the relation $\mathbf{E}^r{}_0=\Gamma^*\mathbf{E}^i{}_0=(\Delta^*-1)\mathbf{E}^i{}_0$, Eq. (11) becomes

$$\mathbf{E}_d=\mathbf{E}^i{}_0(e^{+\gamma d}-e^{-\gamma d}+\Delta^* e^{-\gamma d}) \tag{17}$$

from which it follows that

$$p=\cosh 4\pi\kappa z-\cos 4\pi z+\tfrac{1}{2}(\Delta'^2-2\Delta'+\Delta''^2)e^{-4\pi\kappa z}+\Delta'\cos 4\pi z-\Delta''\sin 4\pi z. \tag{18}$$

This equation has been used for calculating a series of corrections, the results being presented in Table II. The corrections listed for the several values of r' are to be subtracted from the measured PSWR before determining κ graphically from the data in Table I. A simpler alternative to this procedure is to use the nomograph of Fig. 5 to make corrections directly to the value of $\log\kappa$ obtained graphically from the meas-

TABLE II. Short-circuit corrections to power-standing-wave ratios in Table I.

κ	δ db for $r'=0.0025$	δ db for $r'=0.0050$	δ db for $r'=0.0075$	δ db for $r'=0.0100$
0.01	−0.6650	−1.2826	−1.8592	−2.3999
0.02	−0.3353	−0.6649	−0.9750	−1.2824
0.03	−0.2247	−0.4493	−0.6617	−0.8733
0.04	−0.1690	−0.3390	−0.5022	−0.6640
0.05	−0.1356	−0.2718	−0.4036	−0.5350
0.06	−0.1132	−0.2268	−0.3370	−0.4477
0.07	−0.0971	−0.1943	−0.2891	−0.3843
0.08	−0.0850	−0.1698	−0.2531	−0.3365
0.09	−0.0756	−0.1508	−0.2250	−0.2993
0.10	−0.0680	−0.1355	−0.2024	−0.2695
0.11	−0.0618	−0.1230	−0.1839	−0.2448
0.12	−0.0565	−0.1125	−0.1683	−0.2242
0.13	−0.0521	−0.1036	−0.1551	−0.2065
0.14	−0.0482	−0.0960	−0.1436	−0.1913
0.15	−0.0448	−0.0892	−0.1335	−0.1779
0.16	−0.0418	−0.0830	−0.1245	−0.1660
0.17	−0.0391	−0.0778	−0.1166	−0.1554
0.18	−0.0367	−0.0730	−0.1094	−0.1457
0.19	−0.0344	−0.0686	−0.1029	−0.1371
0.20	−0.0323	−0.0646	−0.0969	−0.1292
0.21	−0.0305	−0.0610	−0.0915	−0.1220

ured PSWR. To use the nomograph, a line is drawn through the observed PSWR and the calculated value of r' and extended to intersect the left-hand scale. Corrections made in this manner are commensurate in accuracy with PSWR's of accuracy ± 0.01 db or better. The method for calculating r' from the measured value of R_s/Z_0 is discussed at the end of the following section.

Low Loss Measurements

A convenient modification of the standing-wave method, particularly when the absorption index is of the order of 0.02 or less, is the "breadth-of-the-minimum" technique based upon a procedure which is commonly used for measuring high standing-wave ratios[26] and has been employed by workers in this laboratory for determining complex dielectric constants with waveguide apparatus.[6,7] In this procedure, the breadths of successive minima are measured at some fixed multiple of the minimum power and then plotted as a function of m, the number of each minimum. The curve so obtained increases almost linearly with κ as shown by the following derivation.

As indicated in Fig. 6, the breadth of the mth minimum will be designated as δz_m when measured at a power level which is S times $p^m{}_{\min}$. Assuming that the distance between the mth minimum and its corresponding half-wavelength point is negligible and that the terms $\Delta''^2 e^{-4\pi\kappa z}$ and $\Delta'' \sin 4\pi z$ in Eq. (18) are

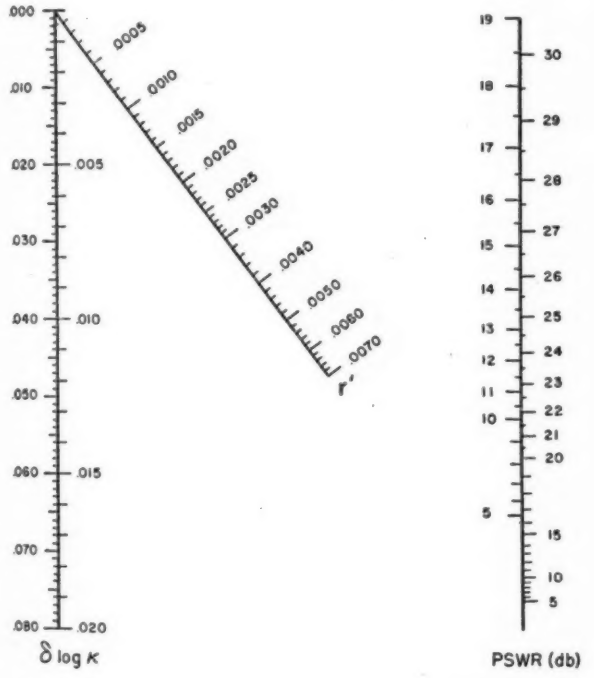


FIG. 5. Nomograph for determining short-circuit corrections in the standing-wave method. (Value of $\delta \log \kappa$ is to be subtracted from the observed value of $\log \kappa$.)

[26] See reference 13, pp. 505, 637.

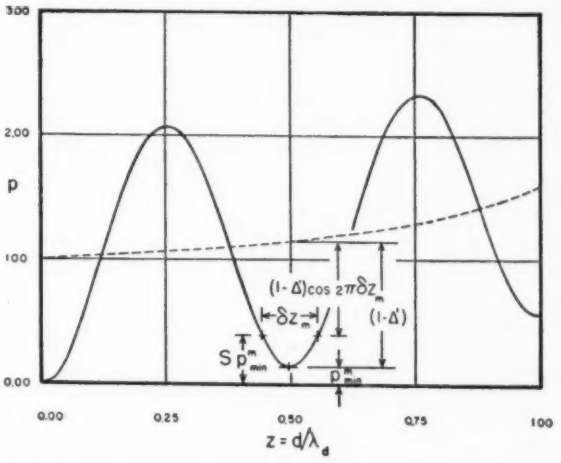


FIG. 6. Power-standing-wave pattern showing definition of the "breadth of the minimum."

insignificant, we may write the approximate equation,

$$p^m{}_{\min} = \cosh 2\pi\kappa m + (\Delta' - 1) + (\tfrac{1}{2}\Delta'^2 - \Delta')e^{-2\pi\kappa m}. \quad (19)$$

To this same approximation, Eq. (18) represents a cosinusoidal variation, of amplitude $(1-\Delta')$, about an axis $\cosh 4\pi\kappa z + (\frac{1}{2}\Delta'^2 + \Delta')e^{-4\pi\kappa z}$, shown as the dashed line in Fig. 6. From this it follows that, if we neglect the change in the axis over the small interval, δz_m,

$$(1-\Delta')(1-\cos 2\pi\delta z_m) = (S-1)p^m{}_{\min}. \quad (20)$$

Finally, by combining Eqs. (19) and (20) and then by expanding in series using Eq. (15), it may be shown that

$$\delta z_m = (S-1)^{\frac{1}{2}}(m\kappa + r'/\pi)(1 + \pi^2 S m^2 \kappa^2/6 + \cdots), \quad (21)$$

or, for most practical purposes,

$$\delta z_m = (S-1)^{\frac{1}{2}}(m\kappa + r'/\pi). \quad (22)$$

The accuracy of Eq. (22) for $\kappa \leqslant 0.02$ is better than 1.2 percent up to $m=3$, while for $\kappa = 0.04$ the error is only 4.8 percent for $m=3$. By using Eq. (21), these errors may be greatly reduced, but the uncertainties in the measurement of δz_m hardly warrant this refinement in most cases.

In practice, the most convenient value of the multiplier S is either 2 or 10; thus, the slope of the correponding δz_m *versus* m curve is either κ or 3κ while the intercept is either r'/π or $3r'/\pi$. When δz_m is being measured, the calibrated piston attenuator is used to reduce the probe signal level by either 3.01 db or 10.00 db relative to $p^m{}_{\min}$, so that the microwave receiver always registers the same output signal just as in the PSWR measurement.

Although the short-circuit resistance does not affect the measurement of κ in the breadth-of-the-minimum method, the finite resistance along the coaxial transmission line is a significant factor when low loss measurements are being made. The necessary correction is best described in terms of the "effective absorption index," κ_0, of the slotted line itself where $\kappa_0 = \alpha_0\lambda_0/2\pi$, and

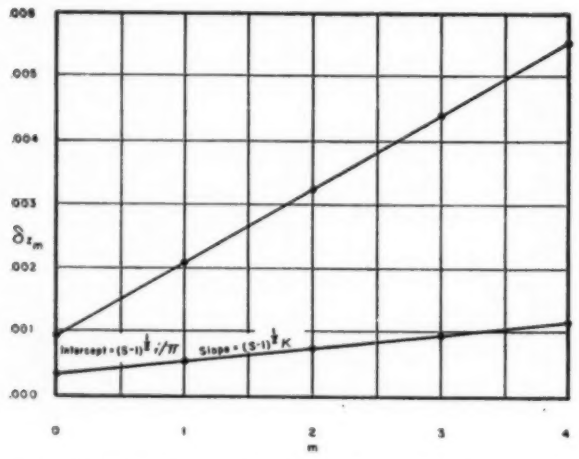


FIG. 7. Plot of breadth of the minimum, δz_m, *versus* m.

α_0 is the attenuation of the empty line. A plot of δz_m *versus* m is made from measurements obtained in the empty slotted line with $S=10$; from the slope of the resulting curve, κ_0 is computed as shown in Fig. 7. The effective absorption index of the slotted line remains unchanged in the presence of a dielectric because while the line attenuation constant increases to $n\alpha_0$, the wavelength decreases to λ_0/n. Therefore, the actual absorption index of the dielectric is obtained simply by subtracting κ_0 from observed value of κ obtained graphically as described above. In the slotted line used by the author, the measured value of κ_0 was 0.00035.

As indicated by Eq. (22), the short-circuit resistance can be computed from the intercept of the δz_m *versus* m curve, this intercept corresponding to the breadth of the minimum at the short circuit. Accordingly, in air, the normalized short-circuit resistance is $r'_0 = \pi\delta z_0/(S-1)^{\frac{1}{2}} = R_s/Z_0$; in the presence of a dielectric, the normalized short-circuit resistance is $r' = n\pi\delta z_0/(S-1)^{\frac{1}{2}} = nr'_0$, δz_0 still being the extrapolated value for breadth of the minimum at the short circuit *in air*. In this manner, the value of r' for use in the nomograph of Fig. 5 may be calculated. The observed value of r'_0 in the author's slotted line was 0.00107 which is considerably greater than the theoretical value,

$$r'_0 = R_s/Z_0 = (\ln b/a)/2\pi\sigma DZ_0 = 1/120\pi\sigma D, \quad (23)$$

where a and b are the inner and outer radii of the coaxial line, σ is the conductivity of the short-circuit material (gold-plated brass), and $D=(1/2\pi)(10^9/\sigma f)^{\frac{1}{2}}$ is the skin depth at frequency f. The cause of this discrepancy was the butt joint between the short-circuiting disk and the outer conductor, this construction being made necessary by the requirement for a removable center conductor.

APPARATUS

The functional diagram of Fig. 1 shows the interrelation of the major components of the apparatus. The signal generator was a 707B or 2K28 klystron oscillator with tunable cavity covering the range from 8.7- to 12.5-cm wavelength. The microwave receiver, having a klystron local oscillator of identical design, was of superheterodyne construction. This permitted continuous-wave reception at a maximum sensitivity level of about 70 db below the full output of the calibrated attenuator. Automatic frequency control was also incorporated in the receiver design and electronically regulated power supplies were used in both signal generator and receiver to provide optimum frequency and amplitude stability. The over-all amplitude stability of the apparatus was 0.1 db, while the frequency stability of the generator was better than 0.01 percent for periods of an hour or less and about 0.1 percent from day to day.

The calibrated piston attenuator was constructed of broached silver tubing of 0.5000±0.0002-in. internal diameter. Thus, with a theoretical attenuation of 15.9905 db per radius at infinite wavelength, the device had an attenuation of almost exactly 64 db per inch. The pick-up loop drive mechanism was linear within ±0.001 in over a distance of 0.75 in., and, since extraneous modes of propagation were reduced to negligible level by the iris coupling, the maximum error of the attenuator was ±0.13 db over a 50-db range.

The piston attenuator in the reference signal channel was of simpler design and was not calibrated. The phase shifter, of the line stretcher type, was likewise uncalibrated, and in order to reduce its amplitude variations it was resistively loaded by a fixed 20-db attenuator.

Since the slotted line presented a bad mis-match when filled with various dielectrics, a double stub tuner, not shown in Fig. 1, was placed at its input and adjusted for maximum power transfer before each measurement. A perfect match was not needed, however, since the method of measurement involves no change of impedance at the input to the slotted line. Thus, the generator sees a constant load throughout the measurement.

The slotted line was machined from brass and then plated with gold to increase its conductivity. Since organic halides were to be measured with the apparatus, gold plating was chosen in favor of silver because of its better corrosion resistance. The center conductor, also made of gold-plated brass, was made removable both to facilitate cleaning and to permit the use of either a short-circuit or a matched termination. Teflon plugs were used both to position the center conductor and to seal against liquids. The entire length of the slotted line, some six inches over-all, was water jacketed to permit thermostating, and a fill pipe was placed near the electrical input end. Liquid capacity of the slotted line was less than 30 ml in keeping with the requirement for small quantities of dielectric material.

The walls of the slotted line were thick enough to permit a slot depth of ¼ in., thus allowing the liquid level to rise to a region of negligible field strength in the slot. Since any discontinuity in the cross section of

the slotted line preceding the short circuit would give rise to an extraneous reflection which would distort the standing-wave pattern, it was necessary to machine the slot flush and square with the plane of the short circuit.

The probe carriage was a narrow slab long enough to keep the slot always covered, thus minimizing contamination and evaporation of the dielectric. The ways for guiding the probe carriage along the slot were ground parallel to the center conductor within 0.0002 in. to insure constant probe depth. Spring-loaded roller bearings kept the probe carriage aligned with the center of the slot, and a micrometer screw of 1-mm pitch, geared to a revolution counter and having an indexed head of 100 divisions, was used to drive the probe carriage along the slot. Probe displacements were discernable to 0.001 mm and accurate to about 0.005 mm.

Instrumental Errors

In the measurement of λ_0 by means of the distance between minima of the standing wave in air, a precision of ± 0.01 percent with comparable accuracy was consistently obtained. In nonpolar (lossless) dielectrics, the wavelength determination was only slightly less accurate.[4] In medium and low loss dielectrics, λ_d was measurable with an accuracy of ± 0.1 percent, corresponding to a ± 0.2 percent error in the computed value of ϵ'. In high loss media, however, the accuracy was somewhat less because of the reduced probe signal amplitude which decreased the null sensitivity.

In the measurement of absorption index, the estimated attenuator error was ± 0.1 db, which corresponds to a *numerical* error of ± 0.0018 in κ, when measured by the traveling-wave method, and to a *percentage* error of ± 0.5 percent in $\log\kappa$ or ± 1.2 percent in κ, when measured by the standing-wave method. In the breadth-of-the-minimum method, the error in κ varied from 2 percent to 10 percent, depending on the magnitude of κ. Thus, for the standing-wave method, the estimated error in the computed value of ϵ'' was 1.4 percent, the errors in the other methods varying with the absorption index.

ACKNOWLEDGMENT

The author wishes to acknowledge the many helpful conversations with his colleagues on this project and their cooperation in making measurements with the apparatus. The assistance of D. M. Olson and L. C. Osborn in preparing the drawings is also gratefully acknowledged.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Propagation of Elastic Waves in a Cylindrical Bore Containing a Fluid

M. A. Biot*
Columbia University, New York, New York
(Received April 7, 1952)

The paper deals with the propagation of unattenuated elastic waves in a cylindrical bore through an elastic material of infinite extent filled with a fluid. The phase and group velocity dispersion curves are plotted for axial symmetric waves in the coupled fluid-solid system.

1. INTRODUCTION

An analysis is here presented of the problem of propagation of elastic waves in a fluid contained in a circular bore through an elastic solid of infinite extent. Only waves of axial symmetry which are pure sinusoids along the axial direction are considered here. We are dealing essentially with the interaction between the compressible fluid and the elastic solid in the generation of propagation phenomena along the axis of the bore.

In view of the complexity of the phenomenon a clearer understanding is attained by first treating in Sec. 2 the problem of propagation of surface waves in the empty bore. In this case the surface wave, which is analogous to a Rayleigh wave at the plane boundary of a semi-infinite solid, exhibits a dispersion. The phase velocity increases with the wavelength from the Rayleigh wave velocity to that of shear waves in the solid. No unattenuated waves are propagated beyond a certain wavelength at which a cutoff occurs. A family of phase and group velocity dispersion curves exists with the Poisson ratio as a parameter.

Propagation in a fluid-filled bore is taken up in Sec. 3. The dispersion curves are first considered for the case of a fluid which is either confined in a rigid bore or suspended in free space. This corresponds to the limiting case of a fluid of zero or infinite density. In this case waves in the fluid are not affected by the elasticity of the solid. The cases of finite density are intermediate between the above, with one exception, and they correspond to interaction between fluid and solid waves. It is pointed out that not all waves result from the inter-

* Consultant Shell Oil Company Exploration and Production Research Laboratory, Houston, Texas. The present paper is based on an unpublished Shell Oil Company report dated October 15, 1949.

action of fluid and solid waves. It is found that one branch of the dispersion curve bears no resemblance to any of the waves occurring independently in either medium but is caused solely by the interaction of the two media. This wave is designated hereafter by the appellation "Stoneley wave" because of its analogy with the waves occurring at the interface of two media. For the larger wavelengths these waves correspond to the well-known phenomenon of the water hammer in tubes. No cutoff occurs for this wave. Plots of phase and group velocity as function of wavelength to diameter ratio are presented for various combinations of the three basic parameters. These parameters are v_s/c, the ratio of shear velocity in the solid to the sound velocity in the fluid, the ratio ρ/ρ_1 of the fluid density to solid density, and Poisson's ratio ν. Work by H. Lamb[1] on this problem appears in a very early publication but results are given only for waves of large wavelengths. As the present analysis shows, this represents only a very limited aspect of the problem.

2. PROPAGATION OF WAVES IN A CYLINDRICAL EMPTY BORE

We consider a cylindrical bore of circular cross section in an elastic solid of infinite extent. This section is concerned with the case where the bore does not contain any fluid. Being a relatively simple case it is treated first, and we focus our attention on the phenomenon of propagation of axial-symmetric waves at the surface of the bore. These waves are the analog of Rayleigh waves propagating at a plane boundary. Our purpose is to determine how the propagation is affected by the axial symmetric nature of the phenomenon. Because a dimension has been introduced, namely, the diameter D of the bore, it must be expected that dispersion will occur; the velocity of the surface wave becomes a function of λ/D, the ratio of the wavelength λ to the diameter.

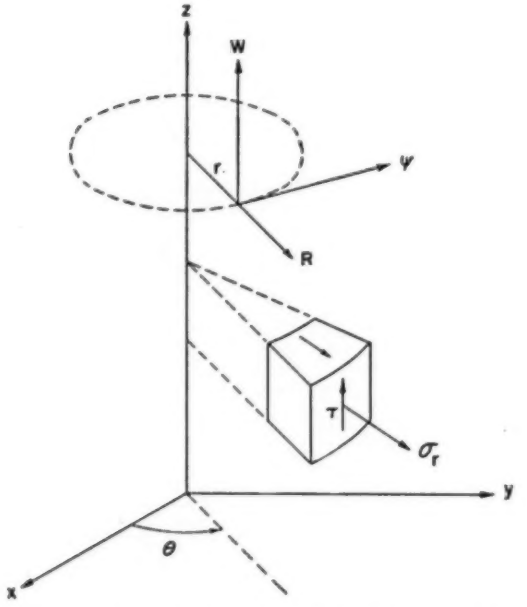


FIG. 1. Cylindrical coordinate system, Z along the axis of the bore.

[1] H. Lamb, *On the Velocity of Sound in a Tube as Affected by the Elasticity of the Walls*, Mem. Proc. Manchester Lt. and Phil. Soc., Vol. XIII, No. 9 (1898).

Since we are dealing with a three-dimensional phenomenon we have three displacement components of the solid, U, V, W. We put

$$(U, V, W)=\text{grad}\phi+\text{curl}\bar{\psi}, \tag{2.1}$$

with

$$\nabla^2\phi=\frac{1}{v_c^2}\frac{\partial^2\phi}{\partial t^2},$$
$$\nabla^2\bar{\psi}=\frac{1}{v_s^2}\frac{\partial^2\bar{\psi}}{\partial t^2}. \tag{2.2}$$

The last equation applies to each Cartesian component of the vector $\bar{\psi}$.

In Eq. (2.2) v_c and v_s represent the velocity of dilatational and rotational waves, respectively.

$$v_c=\left[\frac{2G(1-\nu)}{\rho_1(1-2\nu)}\right]^{\frac{1}{2}},$$
$$v_s=(G/\rho_1)^{\frac{1}{2}},$$

and G=shear modulus of solid, ρ_1=mass density of solid, ν=Poisson's ratio. Because of the axial symmetric nature of the phenomenon, the vector $\bar{\psi}$ is tangent to a circle of radius r, centered on the z axis and parallel with the x, y plane (Fig. 1). Hence

$$\psi_x=-\psi\sin\theta, \quad \psi_y=\psi\cos\theta. \tag{2.3}$$

Similarly, introducing the radial displacement R

$$U=R\cos\theta, \quad V=R\sin\theta. \tag{2.4}$$

The quantities ϕ, ψ, and R functions only of the coordinates r and z. Writing the two Eqs. (2.2) in cylindrical coordinates the first one becomes

$$\frac{\partial^2\phi}{\partial r^2}+\frac{1}{r}\frac{\partial\phi}{\partial r}+\frac{\partial^2\phi}{\partial z^2}=\frac{1}{v_c^2}\frac{\partial^2\phi}{\partial t^2}. \tag{2.5}$$

The second Eq. (2.2), written, for instance, for the component ψ_x is

$$\frac{\partial^2\psi_x}{\partial r^2}+\frac{1}{r}\frac{\partial\psi_x}{\partial r}+\frac{1}{r^2}\frac{\partial^2\psi_x}{\partial\theta^2}+\frac{\partial^2\psi_x}{\partial z^2}=\frac{1}{v_s^2}\frac{\partial^2\psi_x}{\partial t^2}.$$

Introducing the value (2.3) for ψ_x

$$\frac{\partial^2\psi}{\partial r^2}+\frac{1}{r}\frac{\partial\psi}{\partial r}-\frac{\psi}{r^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{1}{v_s^2}\frac{\partial^2\psi}{\partial t^2}. \tag{2.6}$$

The two Eqs. (2.5) and (2.6) for the unknown scalars ϕ and ψ solve the axially symmetric problem.

The displacement components are expressed by means of ϕ and ψ. Equations (2.1) become in the present case

$$R=\frac{\partial\phi}{\partial r}-\frac{\partial\psi}{\partial z},$$
$$W=\frac{\partial\phi}{\partial z}+\frac{\partial\psi}{\partial r}+\frac{\psi}{r}. \tag{2.7}$$

The stress components σ_r and τ (Fig. 1) are

$$\sigma_r=2G\left[\frac{\partial R}{\partial r}+\frac{\nu}{1-2\nu}\epsilon\right],$$
$$\tau=G\left[\frac{\partial R}{\partial z}+\frac{\partial W}{\partial r}\right], \tag{2.8}$$

with

$$\epsilon=\frac{\partial R}{\partial r}+\frac{R}{r}+\frac{\partial W}{\partial z}.$$

The stress components may also be expressed directly in terms of ϕ and ψ. Substituting (2.7) into (2.8) and taking into account (2.5) and (2.6)

$$\sigma_r=\rho_1 b\frac{\partial^2\phi}{\partial t^2}+2v_s^2\rho_1\left[\frac{\partial^2\phi}{\partial r^2}-\frac{\partial^2\psi}{\partial r\partial z}\right],$$
$$\tau=\rho_1\frac{\partial^2\psi}{\partial t^2}+2v_s^2\rho_1\left[\frac{\partial^2\phi}{\partial r\partial z}-\frac{\partial^2\psi}{\partial z^2}\right], \tag{2.9}$$

with

$$b=\nu/1-\nu.$$

The solutions of Eqs. (2.5) and (2.6) for unattenuated waves propagating in the z direction are

$$\phi=\phi_0K_0(mr)\cos(lz-\alpha t)$$
$$\psi=\psi_0K_1(kr)\sin(lz-\alpha t), \tag{2.10}$$

where ϕ_0 and ψ_0 are two constants and

$$m=l(1-\zeta_2^2)^{\frac{1}{2}},\quad \zeta_2=\frac{\alpha}{v_cl}=\frac{v}{v_c},$$
$$k=l(1-\zeta_1^2)^{\frac{1}{2}},\quad \zeta_1=\frac{\alpha}{v_sl}=\frac{v}{v_s}. \tag{2.11}$$

The phase velocity of the waves in the z direction is designated by v. K_0 and K_1 are modified Bessel functions of the second kind of zero and first order. These functions are real and vanish exponentially at infinity. Denoting by $D=2a$ (Fig. 2) the bore diameter, the boundary conditions are

$$\left.\begin{aligned}\tau&=0\\ {}_r&=0\end{aligned}\right\}\quad \text{at}\quad r=a.$$

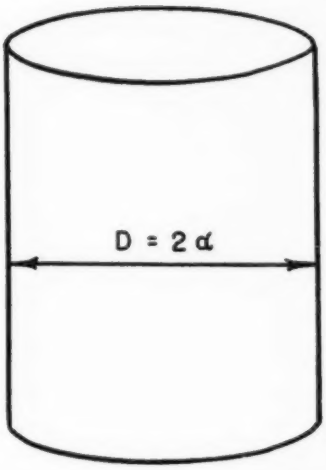


FIG. 2. Portion of cylindrical bore.

The first boundary condition yields a relation between ϕ_0 and ψ_0

$$\psi_0K_1(ka)(2v_s^2l^2-\alpha^2)=2v_s^2ml\phi_0K_0'(ma), \tag{2.12}$$

or

$$\frac{\phi_0}{\psi_0}=\frac{(2-\zeta_1^2)K_1(ka)}{2(1-\zeta_2^2)^{\frac{1}{2}}K_0'(ma)}. \tag{2.13}$$

We use the notation

$$K_0'(u)=\frac{d}{du}K_0(u),$$
$$K_0''(u)=\frac{d^2}{du^2}K_0(u).$$

Introducing the second boundary condition $\sigma_r=0$ we have

$$-\rho_1\alpha^2b\phi_0K_0(ma)+2v_s^2\rho_1[\phi_0m^2K_0''(ma)-\psi_0klK_1'(ka)]=0. \tag{2.14}$$

This expression may be written in terms of K_0 and K_1. We introduce well-known identities satisfied by Bessel functions

$$K_0'(u)=-K_1(u),$$
$$K_0''(u)+\frac{1}{u}K_0'(u)=K_0(u), \tag{2.15}$$
$$K_0''(u)=-K_1'(u)=\frac{1}{u}K_1(u)+K_0(u).$$

With these relations, Eq. (2.14) may be written

$$-b\zeta_1^2\phi_0K_0(ma)+2\phi_0(1-\zeta_2^2)\left[\frac{1}{ma}K_1(ma)+K_0(ma)\right]$$
$$+2\psi_0(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}K_1(ka)+K_0(ka)\right]=0.$$

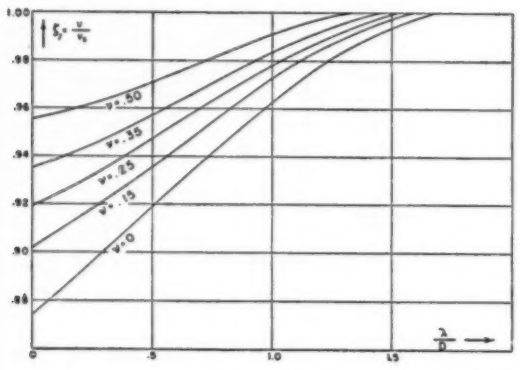


FIG. 3. Phase velocity of surface waves in empty bore.

With the value (2.13) for the ratio ψ_0/ϕ_0 this becomes

$$4(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right]$$

$$-2(2-\zeta_1^2)(1-\zeta_2^2)^{\frac{1}{2}}\left[\frac{1}{ma}+\frac{K_0(ma)}{K_1(ma)}\right]$$

$$+b\frac{\zeta_1^2(2-\zeta_1^2)}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}=0.$$

Substituting

$$b=1-2\left[\frac{\zeta_2}{\zeta_1}\right]^2,$$

we find

$$4(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right]$$

$$-\frac{2(2-\zeta_1^2)(1-\zeta_2^2)^{\frac{1}{2}}}{ma}-\frac{(2-\zeta_1^2)^2}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}=0. \quad (2.16)$$

Since $ma=la(1-\zeta_2^2)^{\frac{1}{2}}$, $ka=la(1-\zeta_1^2)^{\frac{1}{2}}$, this last equation determines the phase velocity of the axial symmetric surface waves as a function of the Poisson ratio and the variable $la=(2\pi a/\lambda)=\pi(D/\lambda)$. Hence the wave shows a dispersion function of the ratio of the wavelength to the diameter of the bore. Physically it must be expected that for very short wavelengths (la large) the velocity must coincide with the velocity of the Rayleigh wave at a plane boundary. This may be verified by putting $la=\infty$ in Eq. (2.16) for large values of the argument the asymptotic value of the Bessel function is

$$K_0(u)=K_1(u)=e^{-u}\left(\frac{\pi}{2u}\right)^{\frac{1}{2}} \quad (u\to\infty).$$

Hence in that case Eq. (2.16) reduces to

$$4(1-\zeta_1^2)^{\frac{1}{2}}-\frac{(2-\zeta_1^2)^2}{(1-\zeta_2^2)^{\frac{1}{2}}}=0 \quad (2.17)$$

which coincides with the equation for the Rayleigh wave at a plane boundary. Equation (2.16) has been solved numerically, and a plot of the velocity ratio $v/v_s=\zeta_1$ as a function of λ/D is shown in Fig. 3 for values of the Poisson ratio ν varying between 0 and $\frac{1}{2}$. A significant feature is that all these curves show an increase of the phase velocity v with increasing wave length until v becomes equal to the value v_s for the shear wave. The curves are terminated at that point which corresponds to a cut-off wavelength λ_c. For wavelengths larger than this wavelength λ_c the waves cannot propagate without attenuation. The ratio λ_c/D depends on the Poisson ratio as follows.

TABLE I.

ν	λ_c/D
0	1.670
0.15	1.583
0.25	1.517
0.35	1.445
0.50	1.310

The curve λ_c/D is plotted against ν in Fig. 4.

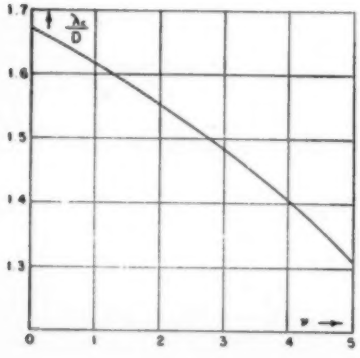


FIG. 4. Cut-off wavelength in empty bore.

The existence of this cut-off wavelength for axial symmetric surface waves is a consequence of two facts: first, that the phase velocity increases with the wavelength, and second, that the radical $(1-\zeta_1^2)^{\frac{1}{2}}$ in Eq. (2.16) must be real. Physically these two facts may be understood by bearing in mind that an axial symmetric deformation gives rise to circumferential stresses which do not exist in the two-dimensional case. This causes an apparent increase of rigidity of the material for such symmetric deformations. This rigidity becomes greater as the wavelength increases, that is, as the depth at which the surface waves are felt becomes larger. It is natural therefore that this apparent increase of rigidity causes an increasing value of the phase velocity. This increase, however, cannot occur beyond the point when the velocity of the surface waves becomes larger than the shear-wave velocity, because a disturbance propagating at the surface with such a velocity causes shear waves to be radiated, and the energy of the surface waves would be dissipated.

Finally it is interesting to derive the group-velocity

curves for such waves. The group velocity v_g is

$$v_g = d\alpha/dl.$$

We may write

$$v_g/v_s = d(\zeta_1 la)/d(la),$$

that is, the ratio v_g/v_s is the derivative of the product $\zeta_1 la$ considered as a function of la. The value v_g/v_s is plotted as a function of $la = \pi D/\lambda$ in Fig. 5 for different values of the Poisson ratio ν.

3. PROPAGATION OF WAVES IN A CYLINDRICAL BORE FILLED WITH A FLUID

We consider the same problem as in the previous section with the addition of a fluid in the bore. We first study the behavior of waves in the fluid independently from the solid. For axially symmetric waves the displacement potential ϕ of the fluid satisfies the wave equation expressed in cylindrical coordinates (c = velocity of sound in the fluid),

$$\frac{\partial^2\phi}{\partial r^2}+\frac{1}{r}\frac{\partial\phi}{\partial r}+\frac{\partial^2\phi}{\partial r^2}=\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}. \tag{3.1}$$

The solution ϕ may be written

$$\phi = J_0\left[r\left(\frac{\alpha^2}{c^2}-l^2\right)^{\frac{1}{2}}\right]e^{i(lz-\alpha t)} \quad \text{for} \quad \frac{\alpha^2}{c^2}>l^2,$$

and (3.2)

$$\phi = I_0\left[r\left(l^2-\frac{\alpha^2}{c^2}\right)^{\frac{1}{2}}\right]e^{i(lz-\alpha t)} \quad \text{for} \quad \frac{\alpha^2}{c^2}<l^2.$$

J_0 is the Bessel function of zero order of the first kind and I_0 the modified Bessel function of zero order of the first kind.

$$J_0(iu) = I_0(u).$$

The first solution corresponds to conical waves which are reflected at the boundary while the second solution corresponds to Stoneley waves.

The fluid pressure for each case is

$$p = -\rho\frac{\partial^2\phi}{\partial t^2} = \rho\alpha^2 J_0[rl(\zeta^2-1)^{\frac{1}{2}}]e^{i(lz-\alpha t)} \quad \text{for} \quad \zeta>1,$$

and (ρ = mass density of the fluid) (3.3)

$$p = \rho\alpha^2 I_0[rl(1-\zeta^2)^{\frac{1}{2}}]e^{i(lz-\alpha t)} \quad \text{for} \quad \zeta<1.$$

We have put $\zeta = v/c$; $v = \alpha/l$ is the phase velocity along the z direction and c is the velocity of sound in the fluid. Denoting the radial displacement of the fluid by

$$R' = \partial\phi/\partial r,$$

and using the relations

$$J_1(u) = -\frac{d}{du}J_0(u) \quad I_1(u) = \frac{d}{du}I_0(u),$$

we may write for each case

$$R' = -l(\zeta^2-1)^{\frac{1}{2}}J_1[rl(\zeta^2-1)^{\frac{1}{2}}]e^{i(lz-\alpha t)},$$
$$R' = l(1-\zeta^2)^{\frac{1}{2}}I_1[rl(1-\zeta^2)^{\frac{1}{2}}]e^{i(lz-\alpha t)}. \tag{3.4}$$

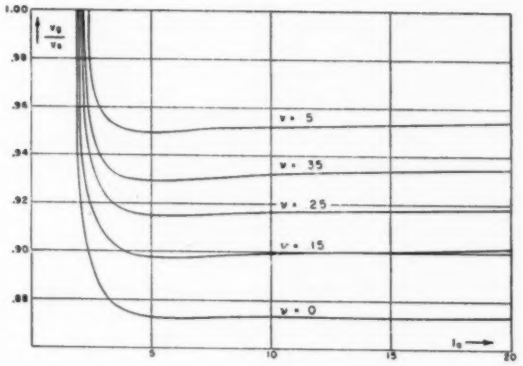


FIG. 5. Group velocity in empty bore.

The corresponding ratios of pressure to fluid displacement at the boundary $r = a$ are therefore

$$\frac{p}{R'} = -\frac{\alpha^2\rho}{l}\frac{J_0[la(\zeta^2-1)^{\frac{1}{2}}]}{(\zeta^2-1)^{\frac{1}{2}}J_1[la(\zeta^2-1)^{\frac{1}{2}}]} \quad \text{for} \quad \zeta>1$$

and (3.5)

$$\frac{p}{R'} = \frac{\alpha^2\rho}{l}\frac{I_0[la(1-\zeta^2)^{\frac{1}{2}}]}{(1-\zeta^2)^{\frac{1}{2}}I_1[la(1-\zeta^2)^{\frac{1}{2}}]} \quad \text{for} \quad \zeta<1.$$

If the fluid is contained in a rigid wall, $R' = 0$ and the corresponding velocities are given by the equation

$$(\zeta^2-1)^{\frac{1}{2}}J_1[la(\zeta^2-1)^{\frac{1}{2}}] = 0. \tag{3.6}$$

The root $\zeta = 1$ corresponds to waves with their plane normal to the axis, while the equation

$$la(\zeta^2-1)^{\frac{1}{2}} = \beta_n \tag{3.7}$$

corresponds to the reflected conical waves. The constants β_n are the roots ($\beta_n \neq 0$) of the Bessel function J_1. For each root, if we plot the velocity ζ as a function of λ/D ($D = 2a$, λ = wavelength) we obtain a hyperbola. Similarly a cylindrical fluid column free at the boundary has reflected conical waves corresponding to the equation

$$J_0[la(\zeta^2-1)] = 0, \tag{3.8}$$

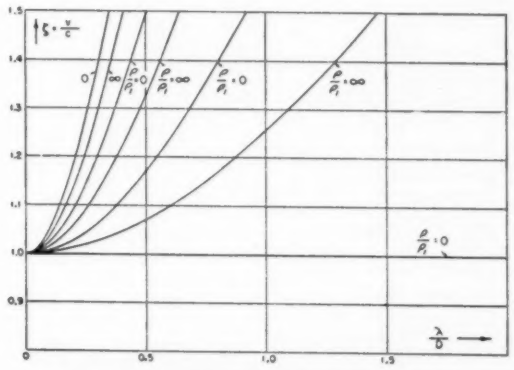


FIG. 6. Phase velocity in fluid for either a rigid wall or the absence of solid wall.

and again this equation represents a family of hyperbolas.

The branches corresponding to Eq. (3.6) are plotted in Fig. 6 as functions of $\lambda/D=\pi/la$. They are marked $\rho/\rho_1=0$ because they also correspond to the case where the wall is elastic but the fluid has a very small density relative to the solid. The branches corresponding to Eq. (3.8) are represented in the same Fig. 6 and marked $\rho/\rho_1=\infty$.

We shall now introduce the effect of the elastic wall. Using equations established in the previous section we derive the ratio of normal stress to displacements at the solid boundary. The stresses in the solid are obtained by substituting expressions (2.10) into (2.9). One of the boundary conditions $\tau=0$ for $r=a$ is the same as in Sec. 2 and leads to the same relations (2.12) and (2.13) between ϕ_0 and ψ_0. For the normal stress σ_r we have at $r=a$

$$\frac{\sigma_r}{2v_s^2\rho_1}=\left[-\frac{b}{2v_s^2}\alpha^2\phi_0K_0(ma)+m^2\phi_0K_0''(ma)-kl\psi_0K_1'(ka)\right]\cos(lz-\alpha t). \quad (3.9)$$

The radial displacement R is given by (2.7)

$$R=[m\phi_0K_0'(ma)-l\psi_0K_1(ka)]\cos(lz-\alpha t).$$

By using Eq. (2.12) we may simplify this to

$$Rl=-\frac{\alpha^2}{2v_s^2}\psi_0K_1(ka)\cos(lz-\alpha t). \quad (3.10)$$

If we divide (3.9) by (3.10) and replace ϕ_0/ψ_0 by its value (2.13), we get

$$\frac{\sigma_r}{v_s^2\rho_1lR}=\frac{b(2-\zeta_1^2)}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_0'(ma)}-\frac{2}{\zeta_1^2}(2-\zeta_1^2)(1-\zeta_2^2)^{\frac{1}{2}}\frac{K_0''(ma)}{K_0'(ma)}+\frac{4}{\zeta_1^2}(1-\zeta_1^2)^{\frac{1}{2}}\frac{K_1'(ka)}{K_1(ka)}. \quad (3.11)$$

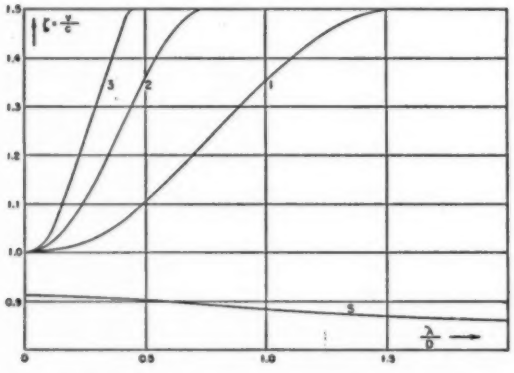


FIG. 7. Phase velocity in fluid-filled bore for $v_s/c=1.5$, $\rho/\rho_1=1$, $\nu=\frac{1}{4}$.

By using identities (2.15) this becomes

$$\frac{\sigma_r}{v_s^2\rho_1lR}=-\frac{b(2-\zeta_1^2)}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}+\frac{2}{\zeta_1^2}(2-\zeta_1^2)(1-\zeta_2^2)^{\frac{1}{2}}\left[\frac{1}{ma}+\frac{K_0(ma)}{K_1(ma)}\right]-\frac{4}{\zeta_1^2}(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right].$$

If we express b as

$$b=1-2[\zeta_2/\zeta_1]^2,$$

we may write

$$\frac{\sigma_r}{v_s^2\rho_1lR}=\frac{(2-\zeta_1^2)^2}{\zeta_1^2(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}+\frac{2}{\zeta_1^2}\frac{(2-\zeta_1^2)(1-\zeta_2^2)^{\frac{1}{2}}}{ma}-\frac{4}{\zeta_1^2}(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right]. \quad (3.12)$$

We now introduce the boundary condition at the fluid-solid interface

$$-p/R'=\sigma_r/R \text{ for } r=a.$$

From (3.5) and (3.12) this may be written for $\zeta>1$ (reflected waves)

$$4(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right]-\frac{2(2-\zeta_1^2)^2(1-\zeta_2^2)^{\frac{1}{2}}}{ma}-\frac{(2-\zeta_1^2)^2}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}=-\frac{\rho}{\rho_1}\frac{\zeta_1^4}{(\zeta^2-1)^{\frac{1}{2}}}\frac{J_0[la(\zeta^2-1)^{\frac{1}{2}}]}{J_1[la(\zeta^2-1)^{\frac{1}{2}}]}. \quad (3.13)$$

Similarly for $\zeta<1$ (Stoneley waves)

$$4(1-\zeta_1^2)^{\frac{1}{2}}\left[\frac{1}{ka}+\frac{K_0(ka)}{K_1(ka)}\right]-\frac{2(2-\zeta_1^2)^2(1-\zeta_2^2)^{\frac{1}{2}}}{ma}-\frac{(2-\zeta_1^2)^2}{(1-\zeta_2^2)^{\frac{1}{2}}}\frac{K_0(ma)}{K_1(ma)}=\frac{\rho}{\rho_1}\frac{\zeta_1^4}{(1-\zeta^2)^{\frac{1}{2}}}\frac{I_0[la(1-\zeta^2)^{\frac{1}{2}}]}{I_1[la(1-\zeta^2)^{\frac{1}{2}}]}. \quad (3.14)$$

The two Eqs. (3.13) and (3.14) determine the dispersion of the waves in the fluid-solid system. The functional relationship of these equations becomes clear if we recall the significance of the symbols. Consider ζ as the unknown variable

$$\zeta=v/c, \quad \zeta_1=c\zeta/v_s, \quad \zeta_2=v_s\zeta_1/v_c.$$

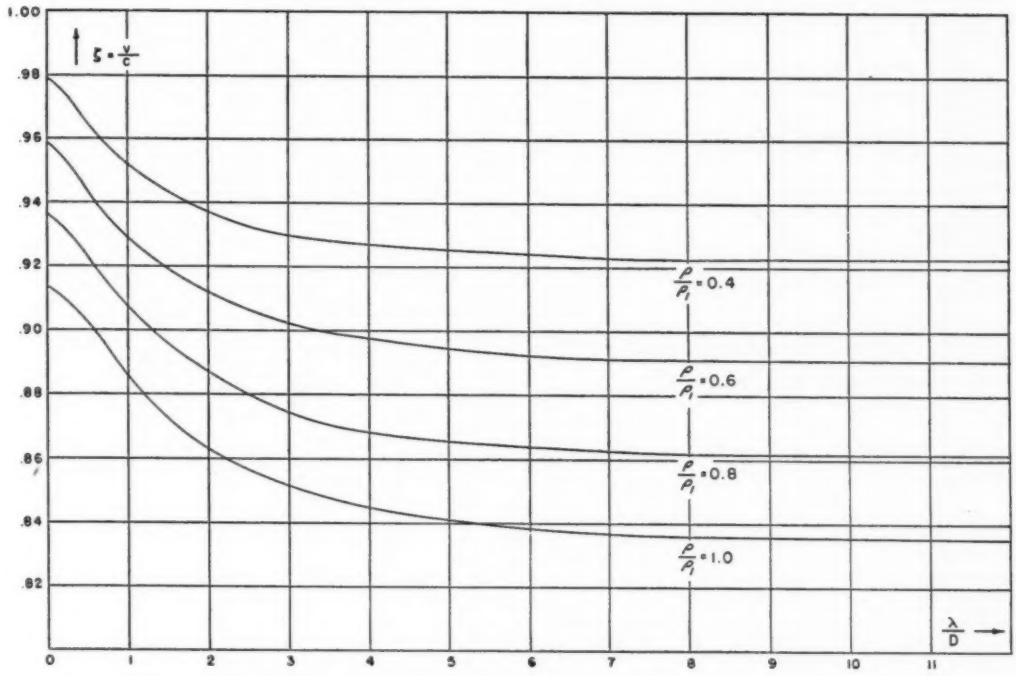


FIG. 8. Phase velocity for Stoneley type wave for various density ratios (branch marked S in Fig. 7).

Hence ζ_1 and ζ_2 are functions of ζ through the parameters v_s/c and v_s/v_c. The parameter v_s/v_c is a function of the Poisson ratio ν only.

$$v_c/v_s = 2(1-\nu)/1-2\nu.$$

When ν varies from zero to 0.5, v_c/v_s varies from $\sqrt{2}=1.414$ to ∞. Also we remember that

$$ka = la(1-\zeta_1^2)^{\frac{1}{2}}, \quad ma = la(1-\zeta_2^2)^{\frac{1}{2}}.$$

Therefore Eq. (3.13) and (3.14) may be considered as determining implicitly the value of $\zeta = v/c$ as a function of the wavelength variable $la = \pi D/\lambda$ for given values of the parameters v_s/c, ρ/ρ_1, ν. These three parameters which characterize the fluid-solid system determine the nondimensional dispersion curves of the waves. Plots have been derived from these equations and will now be discussed.

Figure 7 is a plot of the dispersion curves $\zeta = v/c$ as a function of λ/D(λ = wavelength, D = diameter of the bore) for the following particular values of the characteristic parameters:

$$v_s/c = 1.5, \quad \rho/\rho_1 = 1, \quad \nu = 1/4.$$

These curves show the same qualitative behavior as the curves for waves in the fluid alone, plotted in Fig. 6. The horizontal line of Fig. 6 which corresponded to plane waves becomes branch S in Fig. 7 and represents Stoneley waves propagating with a velocity lower than the velocity of sound in the fluid. The branches marked 1, 2, 3 correspond to the reflected waves. There are an infinite number of such branches and only three of them have been represented.

Figure 8 is a plot of the Stoneley wave branch S of the dispersion curve to exhibit the influence of the density ratio ρ/ρ_1. The value of $\zeta = v/c$ is plotted as a function of λ/D for

$$v_s/c = 1.5, \quad \nu = \tfrac{1}{4},$$

and four values of ρ/ρ_1:

$$\rho/\rho_1 = 0.4,\ 0.6,\ 0.8,\ 1.0.$$

The velocity tends toward a horizontal asymptote for large λ/D. The asymptotic value is found to be

$$\frac{v}{c} = \frac{1}{[1+(\rho c^2/\rho_1 v_s^2)]^{\frac{1}{2}}}$$

or

$$\frac{v}{c} = \frac{1}{[1+(\rho c^2/G)]^{\frac{1}{2}}}. \tag{3.15}$$

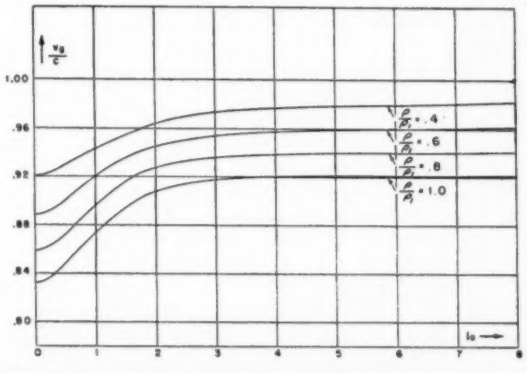


FIG. 9. Group velocity for Stoneley type waves of Fig. 8.

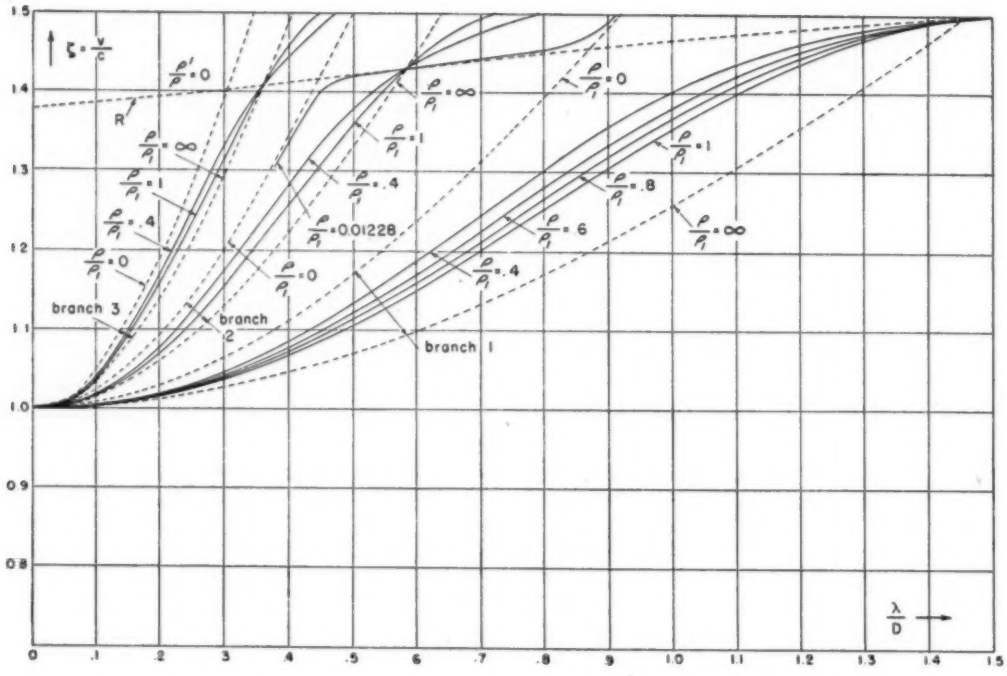


FIG. 10. Phase velocity of reflected type waves for various density ratios (branches marked 1, 2, 3, etc., in Fig. 7).

This formula is valid for all values of ρ/ρ_1 and ν. It shows that the asymptotic value depends only on a single parameter $\rho c^2/G$ which is the ratio of the compression rigidity of the fluid ρc^2 to the shear modulus G of the solid. It will be noted that the phase velocity becomes practically independent of the wavelength λ for ratios λ/D of the order of 5. It may be said that for $\lambda/D > 5$ we are in a region where the Stoneley waves become identical with those studied in the classical theory of the *water-hammer phenomenon*. For small values of the wavelength, the velocity coincides with the Stoneley wave velocity at a plane interface of two semi-infinite media.

Figure 9 is a plot of the group-velocity curves corresponding to the Stoneley waves of Fig. 8. Denoting by v_g the group velocity

$$\frac{v_g}{c} = \frac{d(\zeta la)}{d(la)}.$$

The derivative is taken with respect to $la = \pi D/\lambda$ considered as the independent variable. The value v_g/c is plotted *versus* la for four values of ρ/ρ_1 (density ratio of fluid to solid).

Figure 10 shows a series of phase-velocity curves for the reflected wave branches. The value v/c is plotted *versus* λ/D for the same particular values of the parameters, v_s/c and ν, namely,

$$v_s/c = 1.5, \quad \nu = \tfrac{1}{4}.$$

The first branch is plotted for four values of the

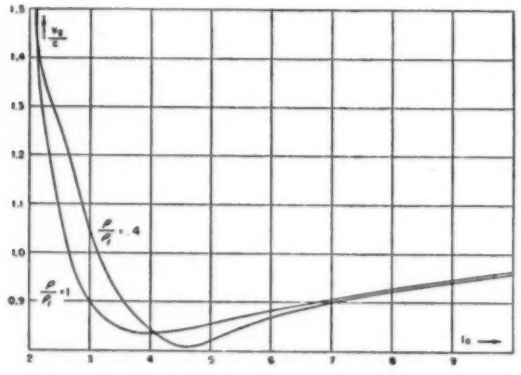


FIG. 11. Group velocity for reflected type wave of branch 1 of Fig. 10 for two density ratios.

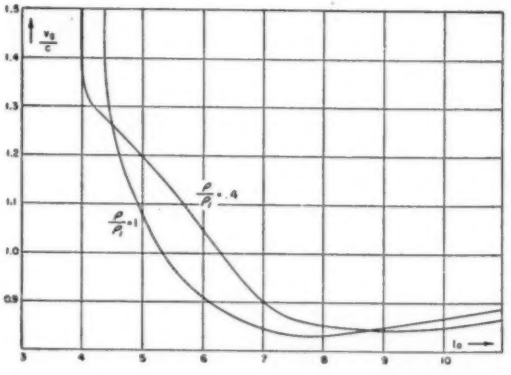


FIG. 12. Group velocity for reflected type waves of branch 2 of Fig. 10 for two density ratios.

parameter

$$\rho/\rho_1 = 0.4,\ 0.6,\ 0.8,\ 1.0.$$

The second and third branches are plotted for two values

$$\rho/\rho_1 = 0.4 \text{ and } 1.0.$$

The second branch is plotted also for the value $\rho/\rho_1 = 0.01228$ in order to show how the curve approaches the case $\rho/\rho_1 = 0$ when ρ/ρ_1 becomes very small. The extreme cases $\rho/\rho_1 = 0$ and $\rho/\rho_1 = \infty$ are plotted as dotted lines. They are made up of the same hyperbolas as in Fig. 6 and of the surface-wave dispersion curve marked R. This surface wave is nothing but a Rayleigh wave of axial symmetry. Its dispersion curve R is the same as the surface-wave dispersion curve for $\nu = \frac{1}{4}$

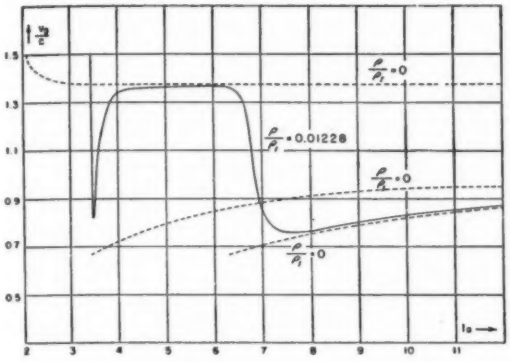


FIG. 13. Group velocity for branch 2 of Fig. 10 for the case of a small fluid density.

plotted in Fig. 3, for the case of an empty bore. The case $\rho/\rho_1 = 0.01228$ illustrates how, for small values of the density ratio, the dispersion branches tend to follow two successive hyperbolas $\rho/\rho_1 = 0$ and the intersecting curve R for the pure surface wave. Whatever the value of ρ/ρ_1 all branches pass through the intersections of the R curve with the hyperbolas $\rho/\rho_1 = \infty$.

Figure 11 shows group-velocity curves for the first branch of the reflected waves of Fig. 10. The ratio v_g/c is plotted *versus la* for the two cases $\rho/\rho_1 = 0.4$ and 1.00. Also $v_s/c = 1.5$, $\nu = \frac{1}{4}$. Note the existence of a minimum value of the group velocity.

Figure 12 shows group-velocity curves for the same cases as in the previous Fig. 11, but for the reflected waves of the second branch.

Figure 13 shows the group-velocity curve for the

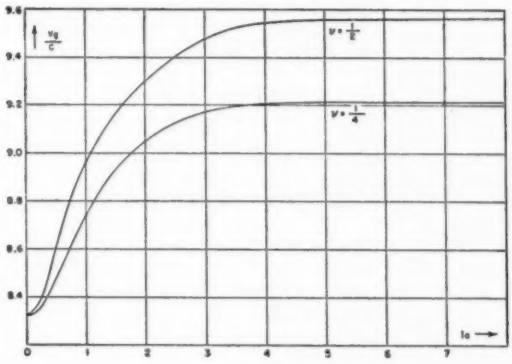


FIG. 14. Influence of Poisson's ratio on the group velocity of the Stoneley type waves.

second branch of Fig. 10 in the case $\rho/\rho_1 = 0.01228$. It is seen that for small values of ρ/ρ_1 the curve tends to follow either one of the three group-velocity curves corresponding to the case $\rho/\rho_1 = 0$ which are shown as dotted lines. The dotted curve on top of the figure is the group velocity for the Rayleigh wave in the empty bore, while the two lower ones are group velocities for the reflected fluid waves between rigid boundaries.

Figure 14 illustrates the influence of the Poisson ratio ν on the group velocity of Stoneley waves for the case $v_s/c = 1.5$, $\rho/\rho_1 = 1$. Two plots of v_g/c *versus la* are shown for $\nu = \frac{1}{2}$ and $\nu = \frac{1}{4}$.

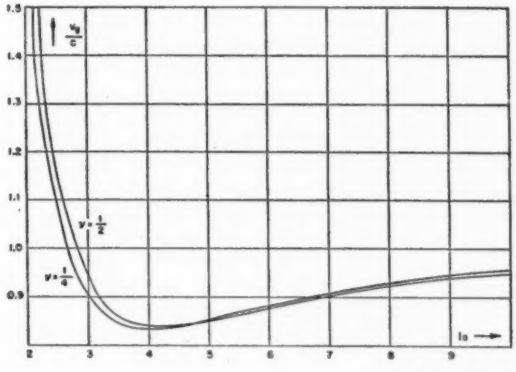


FIG. 15. Influence of Poisson's ratio on the group velocity of reflected type waves of branch 1.

Figure 15 illustrates the influence of the Poisson ratio ν on the group velocity of the first branch of the reflected waves for the case $vs/c = 1.5$, $\rho/\rho_1 = 1$. Two plots of v_g/c *versus la* are shown for $\nu = \frac{1}{2}$ and $\nu = \frac{1}{4}$.

The Conductivity Change in Polyethylene During γ-Irradiation

S. MAYBURG AND W. L. LAWRENCE
Westinghouse Electric Corporation, P.O. Box 1468, Pittsburgh 30, Pennsylvania
(Received May 6, 1952)

The effect of Co^{60} γ-irradiation on the direct current conductivity σ of polyethylene has been determined up to an intensity I of 4000 roentgens per hour at room temperature. The equation

$$\sigma/\sigma_{00}=1+(I/I_0)^{\frac{1}{2}}$$

describes the data where σ_{00} is the conductivity without radiation and is the order of 10^{-18} (ohm cm)$^{-1}$ and I_0 is 20 roentgen/hr. Measurements at liquid nitrogen temperatures give a factor of 20 decrease in the normal conductivity and a factor of 25 decrease in the radiation induced conductivity below their respective room temperature values.

The temperature data serves to suggest an ionic mechanism for electrical conduction in polyethylene as opposed to an electronic mechanism. In fact, the ion contributing most to the conduction in the polyethylene may well be the proton.

I. INTRODUCTION

A PREVIOUS study of the effect of γ-irradiation on the electrical conductivity of three good insulators was made by Armstead, Pennock, and Mead[1] at the Massachusetts Institute of Technology where they irradiated Okoseal, Saran, and polystyrene with 2.5-Mev x-radiation. The conductivity of all three materials increased and reached an equilibrium value under irradiation which depended on the radiation intensity. When the radiation was removed, the conductivity of the three materials decreased to their initial values. In particular it was found that the conductivity of Okoseal increased by a factor of 30 with an exposure intensity of 400 roentgen/min, the conductivity of Saran increased by a factor of 3 with 2000 roentgen/min, and the conductivity of polystyrene increased by a factor of 2×10^3 with 100 roentgen/min. It is interesting to note in this connection that tests on the permanent change in the conductivity of polystyrene exposed in the Oak Ridge Reactor[2] indicate no change up to an integrated exposure of 9×10^{18} thermal neutrons/cm^2, which corresponds to roughly 10^{10} roentgens of γ-exposure. The case of polystyrene clearly indicates that the instantaneous changes of the conductivity of an electrical insulation during γ-irradiation may be far more important than the long term chemical deterioration effects of the irradiation.

Armstead, Pennock, and Mead give no data for the exact dependence of the equilibrium conductivity on the intensity of the γ-irradiation for the various materials. For engineering applications and in order to attempt to understand the mechanism for electrical conduction in good insulators during irradiation, such data are necessary. We have selected polyethylene as the insulation to be used for the study of the dependence of electrical conductivity on intensity because it is readily available in long lengths of coaxial cable. In this form it is easily used in a hot laboratory. Also polyethylene has good long term stability against radiation damage[3] so that all changes in the conductivity produced by irradiation can be assumed to be independent of previous irradiations.

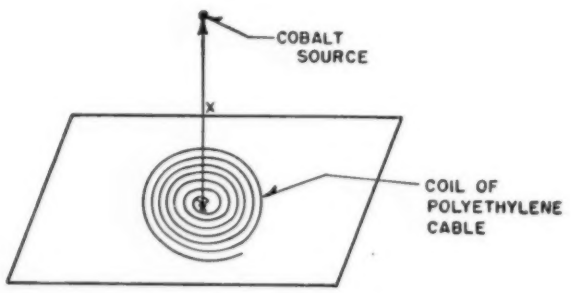


FIG. 1. Irradiation geometry.

II. EXPERIMENT

A. Sources

The large neutron fluxes available in the various piles has allowed the production of larger quantities of radioactive isotopes than were obtainable heretofor. Such an isotope, extremely helpful as a γ-ray source, is Co^{60}. This isotope has a 5.3-year half-life and yields on decay 2 photons, one of energy 1.1 Mev and one of energy 1.3 Mev.[4] Because of the long half-life, the intensity of cobalt 60 sources need only be calibrated once in the course of an experiment providing the experiment does not extend longer than two or three months. Calibration of the cobalt sources was made by comparison with a standard radium source. Because of the large activities which can be induced in Co^{60} by pile irradiation, small amounts of the material can serve as "point" sources and still yield appreciable γ-ray intensities.

The unit of γ-ray intensity, the roentgen, has been carried over from x-ray work. The roentgen is that amount of γ-radiation which will, on passing through pure air at STP, produce one esu of positive and one esu of negative ions per cc. In estimating the γ-ray intensity of a cobalt source we shall use the following relation between photoflux and intensity in roentgens

[1] Armstead, Pennock, and Mead, Phys. Rev. **76**, 860 (1949).
[2] O. Sisman and C. D. Bopp, ORNL-928 (June 29, 1951), p. 163.
[3] See reference 2, p. 78.
[4] L. D. Marinelli, Nucleonics **8**, S 21, (1951).

per hour for a photon of 1.2 Mev:

$$1 \text{ photon/cm}^2 \text{ sec} = 2.2 \times 10^{-6} \text{ roentgen/hr.}$$

B. Experimental Set-Up

All measurements were made inside of a hot laboratory cell having walls and windows thick enough to provide shielding of personnel from the cobalt γ-rays. The polyethylene cable to be irradiated was formed into a flat coil as shown in Fig. 1. The cable used was Amphenol RG 59 U. The source was placed at various positions x along the axis of the coil. Such a geometry has the advantage that no piece of cable shields any other and the γ-intensity at any point of the coil can be easily calculated. Back scattering from the cable support is unimportant for γ-energies as high as 1.1 Mev.

Approximately 75 feet of cable were so coiled, and an additional 30 feet were used as leads into the shielded cell. A cobalt source of activity 8.6 curies was placed at 30, 24, 18, 12, 6, 3, and 0 in. from the center of the coil along its axis. In all but the nearest position the cobalt could be considered as a point source. The direct current resistance was measured between the center and outer conductor of the coaxial cable.

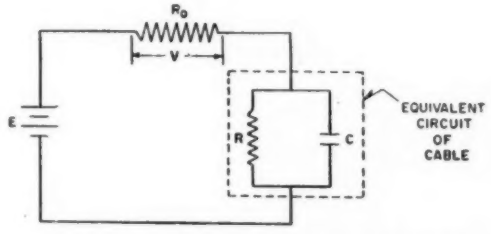


FIG. 2. Circuit for resistance measurement.

A Beckman ultrohmeter was used for these measurements. The absolute accuracy of the ultrohmeter is 5 percent. However, relative measurements are good to 1 percent. The accuracy of measurements of the intensity dependence of the conductivity is determined by the relative accuracy of the instrument. The circuit for measuring the cable resistance R is shown schematically in Fig. 2. The cable itself is represented by the resistance R and capacitance C in parallel. The current flowing through the standard resistance R_0 is determined by measuring the voltage drop (V) across it. After a sufficient time of measurement $>R_0C$ the current in the cable will flow primarily through the resistance R, and hence R can be determined as

$$R=\frac{E-V}{V/R_0}=(E/V-1)R_0.$$

Measurements of R *vs* position of the cobalt source was taken for $E=20$, 91.5, and 580 volts. These measurements are plotted in Fig. 3. The symbol R_{00} denotes the resistance measured with the source at infinity, i.e., no

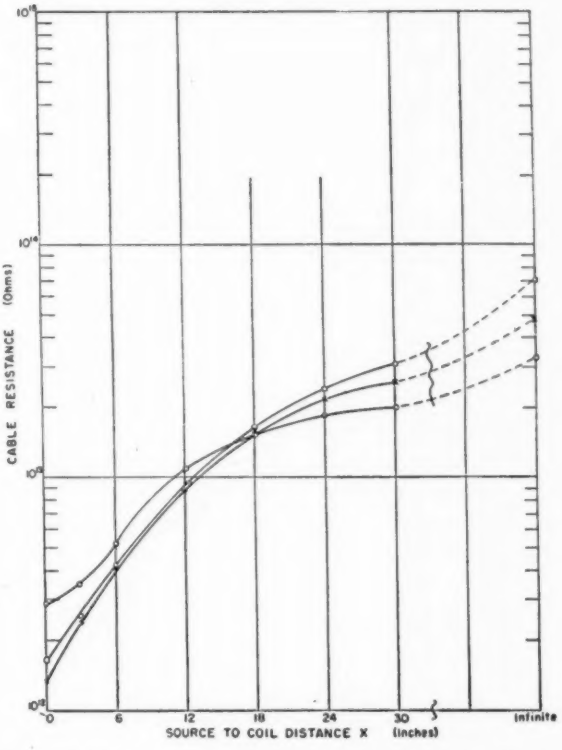


FIG. 3. Initial resistance data. The source is 8.6 curies; the circles represent data taken at 20 volts, the crosses data at 91.5 volts, and the squares data at 580 volts. The cable resistance is plotted as a function of the distance between the source and the coil.

irradiation. Note that in this geometry, resistance changes of as much as a factor of 50 were obtained by the irradiation. The different shape of the 20-volt curve is not significant because of larger possible errors in this data. These errors are discussed in the next section.

C. Analysis of Data

Because of the particular cable geometry employed, it is necessary to make a general assumption regarding the functional form of the dependence of the conductivity on intensity in order to analyze the data. The assumption is made that

$$\sigma(I)=\sigma_{00}+\sigma_1(I)\equiv\sigma_{00}(1+cI^n), \qquad (1)$$

where σ_{00} is the conductivity of the cable without radiation, I is the radiation intensity, and c and n are constants to be determined. This assumption allows for the subtraction of the conductance of the leads into the shielded cell and of any leakage conduction outside the measuring instrument. A new resistance R_1 may be defined in terms of R_{00}, the resistance of the coil measured with no radiation intensity, and R, the resistance measured with irradiation. The relation is

$$1/R_1=1/R-1/R_{00}.$$

Now, R_1 may be related to σ_1 by the usual law that the

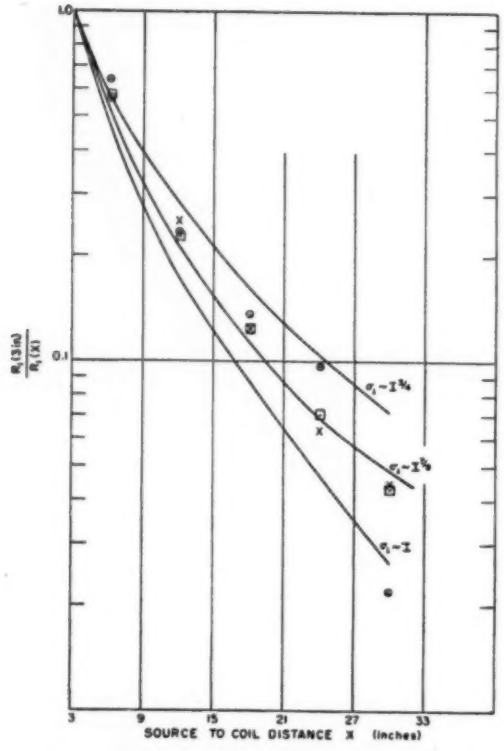


FIG. 4. The ratio of the radiation resistance R_1 (3 inches) to $R_1(x)$ as a function of x, the source to coil distance. The source is 8.6 curies. The circles represent data taken at 20 volts, the crosses data at 91.5 volts, and the squares data at 580 volts. The solid lines are the theoretical curves for $\sigma_1 \sim I^{3/4}$, $I^{7/8}$, and I.

conductivity of resistances in parallel is additive. That is

$$1/R_1 \sim \int \sigma_1(I)dy,$$

where y is the length of the cable.

To find dy, we note that at a radial distance r from the axis of the coil there is a length of cable $2\pi r$ and in the region between r and $r+dr$ there are dr/t such lengths where t is the cable thickness. Hence, $dy=2\pi r(dr/t)$ and

$$1/R_1 \sim \int \sigma_1 2\pi r(dr/t).$$

Since $I \sim (x^2+r^2)^{-1}$, the dependence of R_1 on x can be calculated as follows. Assume n in Eq. (1) is 1. Then

$$1/R_1 \sim \int_0^a \frac{rdr}{x^2+r^2},$$

where a is the radius of the coil (8 in. in the present case). This yields

$$1/R_1 \sim \log(1+x^2)/a^2.$$

Assuming $n=\frac{3}{4}$, we find

$$1/R_1 \sim (x^2+a^2)^{1/4}-\sqrt{x},$$

and $n=\frac{7}{8}$ gives

$$1/R_1 \sim (x^2+a^2)^{1/8}-x^{1/4}.$$

In Fig. 4, the ratio of R_1 measured at $x=3$ in. to R_1 measured at x is plotted as a function of x. The point $x=3$ in. is taken as reference because it is the closest point where the cobalt can be considered a point source. In the same figure are plotted the theoretical curves derived from assuming $n=\frac{3}{4}$, $\frac{7}{8}$, and 1. A comparison of the points derived from the experiments with the curves leads one to conclude that n in Eq. (1) is in the range $\frac{3}{4}$ to $\frac{7}{8}$. The larger scatter in the points of the 20-volt data in Fig. 4 arises from the large error in measuring small currents with the ultrohmeter. Also, any space charge induced voltages in the polyethylene by the ionizing effects of the radiation will introduce a larger error at low voltages than at high voltages. Further, the results of the experiments conducted at 91.5 volts and 580 volts as shown in Fig. 4 suggest that n is voltage insensitive in this region of voltage.

The assumption expressed by the equation

$$\sigma(I)=\sigma_{00}+\sigma_1(I)$$

was necessary in order to subtract away the normal conductance of the cable from the conductance induced by radiation. However, this assumption may be bypassed in determining the dependence of conductivity on radiation if sufficiently high intensity sources are available. Therefore, the dependence of the conductivity on the intensity was redetermined when a 33-curie source of cobalt became available.

In Fig. 5, the relative conductances obtained using the 33-curie source are plotted in a manner analogous to that of Fig. 4.

The experimental points with the 33-curie source leaves no doubt that the conductivity at high intensity is given by the equation

$$\sigma \sim I^{3/4}.$$

Thus the conductivity of polyethylene is proportional to the three-fourths power of the intensity. For the whole intensity region one may write Eq. (1) in the form

$$\sigma/\sigma_{00}=1+(I/I_0)^{3/4}. \qquad (1a)$$

From the data, an estimate of I_0 can be made. We find that, to within a factor of 2,

$$I_0 \sim 20 \text{ roentgen/hr.}$$

TABLE I.

	No radiation	Radiation
Room temperature (300°K)	σ_0	$4.2\times10^2\sigma_2$
Liquid nitrogen temperature (77°K)	$1/20\sigma_0$	$1/25(4.2\times10^2)\sigma_0$

Formula (1a) represents the data up to 4000 roentgen/hr, which was the highest intensity used in these experiments.

D. Effect of Temperature on the Radiation Induced Conductivity

Information on the temperature dependence of conductivity is usually suggestive of the mechanism of conduction. Therefore, temperature measurements were made on polyethylene with and without radiation. About 80 feet of polyethylene cable was coiled in a Dewar flask which was placed in a hot cell. Twenty feet of cable was used as leads into the cell. Table I presents the data on the direct current conductance of the polyethylene cable at room and liquid nitrogen temperatures and with and without radiation. The radiation intensity for these measurements was not known absolutely. However, the radiation conditions were identical at the two temperatures of measurement. The data in Table I includes corrections for the conductance of the leads into the hot cell and the different γ-ray absorption of air and of liquid nitrogen in the Dewar flask. Sufficient errors still exist in these measurements to make it hard to say that there is a real difference between the factor 1/20 in the first column of Table I and the factor 1/25 in the second column of Table I. Therefore, we may conclude that the effect of temperature on the normal and radiative conductivity is ths same.

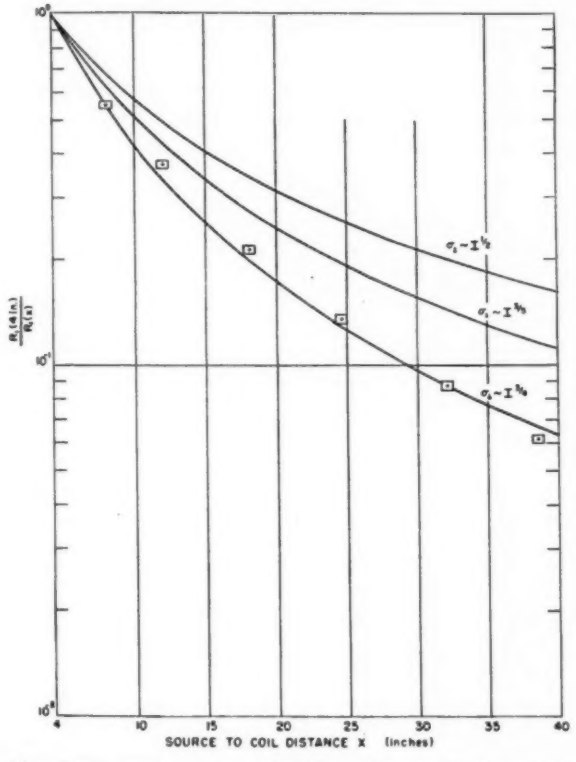


FIG. 5. The ratio of the radiative resistance R_1 (4 inches) to $R_1(x)$ as a function of x, the source to coil distance for a source of 33 curies. The solid lines are the theoretical curves for $\sigma_1 \sim I^{1/2}$, $I^{3/5}$ and $I^{3/4}$.

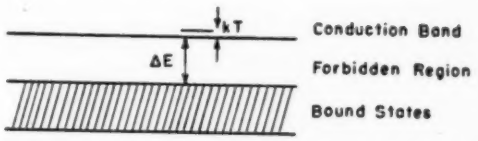


FIG. 6. The band picture for a free electron in an insulator.

III. THEORY

The conductivity of polyethylene, $(CH_2)_n$, during bombardment with high energy γ-rays has many similarities with the low energy photon induced conductivity in insulating crystals. The only obvious difference between γ-conductivity and photoconductivity is that a γ-ray may free many more electrons in a given material than a photon. Extensive work on photoconductivity shows that the photocurrent or -conductivity during irradiation depends on the radiation intensity I by the relation[5]

$$\sigma \sim I^{1/2}.$$

In the measurements reported here on the intensity dependence of the γ-conductivity of polyethylene we found instead of the $I^{1/2}$ law

$$\sigma \sim I^{3/4}.$$

This discrepancy between the observed intensity dependence of the γ-conductivity and that predicted from our knowledge of photoconductivity raises the doubt whether the conduction in polyethylene is by electrons as it is in the photoconductive crystals. Conceivably, a mobile ion such as the proton could be the major charge carrier in the polyethylene.

The solution to the problem of whether conduction is by ions or by electrons can be best given by studies of the temperature dependence of the equilibrium conductivity during irradiation and of the decay at different temperatures of the conductivity when the radiation is removed. Let us explore the possibilities of the former type of study for determining the conduction mechanism.

The conductivity σ is given by the relation

$$\sigma = ne\mu,$$

where n is the number density of charge carriers, e their charge, and μ their mobility. Consider first that electrons are the charge carriers. During irradiation the number of free electrons n should be highly temperature independent because the recombination rate p of an electron with a positive ion is nearly temperature independent. To see this consider the band picture in Fig. 6. The probability per second that a free electron in the conduction band drop down to a bound state (that is, reunite with a positive ion) will be primarily a function

[5] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1948), p. 123.

of the energy difference between the initial and final states. In other words,

$$p=f(\Delta E+kT).$$

Now, for any good insulator $\Delta E \ll kT$. For example, ΔE for diamond is 6 ev and for the semiconductor silicon is 1.1 ev, while kT at room temperature is 0.025 ev. Therefore, p may be expanded as

$$p=f(\Delta E)+kTf'(\Delta E).$$

For any reasonable function f, p is very insensitive to temperature changes.

The mobility μ of an electron will change with temperature according to the relation of Pearson and Bardeen,[6]

$$1/\mu=a/T^{3/2}+bT^{3/2},$$

where when $b=0$ scattering by impurity atoms dominates and when $a=0$ scattering by thermal deviations of the lattice from a perfectly periodic potential dominates. In any case the maximum factor by which the mobility of an electron changes between two temperatures T_1 and T_2 is given by

$$(T_2/T_1)^{3/2}.$$

Between room temperature (300°K) and liquid nitrogen temperature (77°K) this maximum change in mobility is a factor of 7. In Table I data are presented which show that the radiation induced conductivity changes by a factor of 25 between room and liquid nitrogen temperatures. On the assumption of electronic conductivity we can only account for a change by a factor of seven.

Let us now consider the possibility of ionic conduction. In particular, let us suppose that only the most mobile ion in the polyethylene, the proton, is doing the conduction.

Then

$$\sigma=Ne\mu,$$

where for ionic conduction not only the mobility μ will change with temperature but also the equilibrium number N of ions or protons at a given intensity will change with temperature. The number N depends on temperature because the recombination of ions is temperature dependent. In fact the recombination rate R may be expressed as

$$R\sim N^{\alpha}/\tau,$$

where τ is the mean time it takes a proton to unite with a negative ion and α is a number which would be 2 in the absence of a long-range attractive force between the proton and the negative ion. Actually we can determine α for polyethylene from the measurements of the dependence of conductivity on intensity by equating the ion recombination rate R to the rate of production of ions P. Since P is proportional to the intensity I we have

$$R=P,$$
$$N^{\alpha}1/\tau\sim I,$$

or

$$N\sim(\tau I)^{1/\alpha}.$$

But since the conductivity

$$\sigma=Ne\mu,$$

then

$$\sigma\sim(\tau I)^{1/\alpha}\mu. \tag{2}$$

In order to get the observed dependence of σ on $I^{3/4}$ we must set

$$\alpha=4/3.$$

Now, τ depends inversely on the velocity of approach of the ions which are recombining. Since the velocity is proportional to the mobility, we have

$$\tau\sim 1/\mu$$

or

$$\sigma\sim I^{3/4}\mu^{1/4}. \tag{3}$$

The temperature dependence of the conductivity during irradiation is contained in the temperature dependence of the ionic mobility μ. Notice that since the mobility of an ion increases with temperature the conductivity should increase with temperature. As we see in Table I, the conductivity does in fact increase with temperature.

The ionic mobility μ may be related to the ionic diffusion constant D by the Einstein equation[7]

$$\mu=(e/kT)D.$$

Ionic diffusion is a "rate" process governed by an activation energy, and the diffusion constant may be expressed as

$$D=D_0e^{-Q/kT},$$

where D_0 is temperature independent and Q is the activation energy for diffusion of a proton. Equation 3 may now be rewritten

$$\sigma_{\text{ionic}}\sim T^{1/4}e^{-(Q/4kT)}.$$

Using the data presented in the second column of Table I, we find for Q

$$Q=0.13 \text{ ev}.$$

This is a very low activation energy. However, it is not unreasonable in view of the openness of the polyethylene structure, the low mass of the proton, and its lack of orbital electrons. In fact, studies of the diffusion of hydrogen in metals give activation energies of this same order of magnitude.[8]

The above arguments do not constitute a proof of ionic conduction in polyethylene; rather they indicate that it may be possible to supply a proof if sufficient

[6] G. L. Pearson and J. Bardeen, Phys. Rev. **75**, 865 (1949).

[7] See reference 5, p. 63.

[8] R. M. Barrer, *Diffusion in and Through Metals*, (Macmillan Company, New York, 1941), pp. 220–223; R. H. Fowler and G. J. Smithells, Roy. Soc. Proc. (London) **160A**, 43 (1937).

conductivity measurements are made during irradiation at different temperatures. With sufficient points the characteristic $\log\sigma vs 1/T$ curve could be obtained. If this curve were a straight line, we would have proof of an activation process and hence of ionic conduction.

IV. CONCLUSION

The conductivity of polyethylene during γ-irradiation is found to depend on the three-fourths power of the radiation intensity instead of the one-half power as might be expected from an analogy with photoconductivity. Preliminary temperature measurements during irradiation indicate that ions may be doing the conduction in the polyethylene.

These results indicate that further experiments of the following type may be well worthwhile.

1. More detailed studies of the dependence of the radiation induced conductivity on temperature.
2. Quantitative studies of the rate of build-up and decay of the conductivity when the irradiation is applied and removed at different temperatures. Such experiments may help to clarify the phenomenon known as dielectric absorption.
3. Studies of other insulators besides polyethylene, especially those containing no hydrogen such as Teflon $(CF_2)n$. Activation energies for the motion of carbon or of fluorine ions in Teflon should be much higher than for protons in polyethylene..

ACKNOWLEDGMENT

The authors are pleased to acknowledge the aid of Dr. R. T. Jones and Dr. K. Shure in making available the hot laboratory facilities. In addition, our colleagues of the Solid State Physics Group, which is under the direction of Dr. W. E. Johnson, have offered many helpful suggestions during the course of the work.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Metal Transfer and Wear*

I.-MING FENG
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 16, 1952)

A new theory of metal transfer and wear is suggested in this paper. The essence of the theory is as follows: Metal transfer and wear take place at points of actual contact. The interfaces of the high spots that actually make contact are roughened as the result of plastic deformation when they carry normal load. The mechanical interlocking effect of these roughened interfaces is the primary cause of metal transfer and wear. Due to the mechanical interlocking effect of the roughened interface, and the strain-hardening that accompanies plastic deformation, the application of a tangential force will break one of the pair of the contacting high spots a certain distance away from the interface rather than at the original interface. A secondary cause of metal transfer is the adhesion or the diffusion process which takes place during the temperature flash that occurs during breakage. If the adhesive force is very weak and the diffusion process is not rapid enough to cause the sheared-off peak of the high spot to become a blob of transferred metal, the small piece of metal sheared from the high spot can leave as a loose wear particle. This proposed theory explains the welding of the sheared-off peak to its opponent high spot as the consequence of friction, whereas, in the "welding" theory of friction, welding is considered as the cause of friction. Most metallic surfaces in ordinary atmosphere are always covered by a surface film. The effect of surface contamination on metal transfer, wear, and the shear component of friction is discussed. Difficulties encountered in applying the "welding" theory to explain the friction of and metal transfer between contaminated surfaces where metallic adhesion is absent, are obvious. Experimental support of this new theory is given here. It includes as direct evidence (1) the roughening of the interface as the result of the plastic deformation, (2) the perfect matching at the roughened interface, which gives a strong mechanical interlocking effect, and (3) a definite region of severely strain-hardened metal near the interface.

INTRODUCTION

BOTH metal transfer and wear produce a serious effect on moving parts. The evidence of a transferred metallic blob has been beautifully demonstrated by Bowden, Moore, and Tabor[1] employing the taper section technique, as well as by Sakmann, Burwell, and Irvine[2] with the aid of the radioactive tracer method. Unfortunately, a basic understanding of the fundamental nature of these phenomena has been lacking.

In this paper, the author presents a newly developed theory of metal transfer and wear and experimental support. The detail analysis of the effect of various factors on metal transfer and wear and the formulation of the relation among the factors and the amount of metal transfer and wear will be presented in later publications.

THE AREA OF ACTUAL CONTACT AND THE SURFACE TEMPERATURE FLASH

Many interesting experimental results in the field of friction have been published, among which the following are of great importance.

The experimental results of Holm and his collab-

* Containing a part of the dissertation submitted by the author in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Michigan.

[1] Bowden, Moore, and Tabor, J. Appl. Phys. **14**, 80–91, (1943).

[2] Sakmann, Burwell, and Irvine, J. Appl. Phys. **15**, 459–73, (1944).

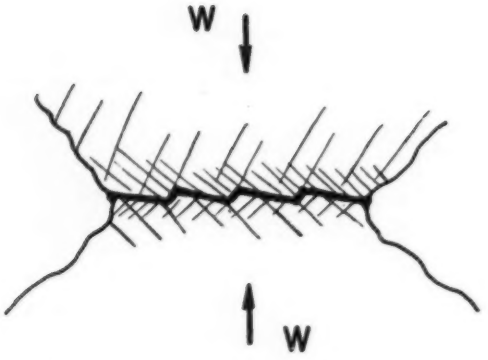


FIG. 1(a). Sketch showing the roughened interface of a pair of actually contacting high spots as the result of plastic deformation.

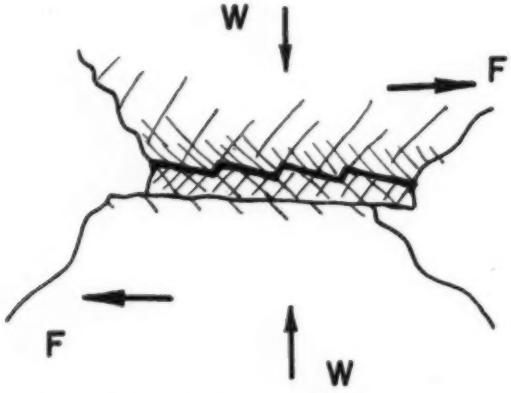


FIG. 1(b). Sketch showing that the mechanical interlocking effect of the roughened interface offers a great resistance to the tangential force and that the breakage occurs along the weakest section rather than the original interface.

orators[3] have shown that the area of actual contact between two contacting surfaces is only a very small portion of the total apparent area. This suggests that solid surfaces consist of numerous irregularities and that, when two surfaces are brought together, a great many scattered high spots, instead of the entire apparent area, come into contact. Consequently, the bearing pressure at most of the actually contacting high spots will be so high that they are plastically deformed.

Bowden and Leben[4] have reported the observation of a surface temperature flash during their dry friction experiments. The surface temperature flash has also been observed by Morgan, Muskat, and Reed.[5]

METAL TRANSFER AND WEAR

The above cited experimental results have led the author to suggest the following theory of metal transfer and wear.

Metal transfer and wear take place at the plastically deformed high spots of actual contact. The interface of a pair of plastically deformed high spots of actual contact, as shown in Fig. 1(a), is roughened as the result of plastic deformation. The roughened interface of the contacting high spots gives the mechanical interlocking effect which is the primary cause of metal transfer and wear. The mechanical interlocking effect of the roughened interface offers a great resistance to the tangential force and, therefore, will resist any relative motion along the interface in the tangential direction. Near the interface, the metal has been severely strain-hardened. The strain-hardening effect falls off rapidly, and the cross section of the high spot increases, as the distance from the interface increases. The application of a tangential force to produce relative motion between two surfaces will break one of the pair of contacting high spots at a certain distance away from the interface where the section is weakest, as shown in Fig. 1(b). The liberation of heat during the process of breaking the high spot causes a sudden, rapid rise of temperature followed by a rapid fall of temperature. When the heat liberated during the breakage of the high spot is conducted through the interface, the temperature flash causes a momentarily rapid diffusion at the roughened interface, which is the secondary cause of metal transfer. The sheared-off peak of the broken high spot may become a piece of transferred metal on the opponent high spot if the diffusion process is rapid enough during the temperature flash to cause welding. If, however, the diffusion process is not rapid enough to cause welding, the adhesive force may be sufficient to keep the sheared-off peak on the opponent high spot as a piece of transferred metal. In case the adhesive force is very weak and the diffusion process is not rapid enough to cause welding, the small piece of metal sheared from the high spot drops off as a loose wear particle.

The numerous high spots on a "flat" surface are of various sizes and heights. When two "flat" surfaces are brought together, some high spots come into contact first. The application of a normal load will push these contacting high spots together and, at the same time, bring new high spots into contact. Because of the statistical distribution of the size and height of the high spots, the contacting high spots are by no means carrying equal load, nor are they deformed to the same extent. Some are only elastically deformed and, of course, carry a very small load; some are plastically deformed. The elastically deformed high spots of actual contact contribute nothing to metal transfer and wear. Among the plastically deformed high spots, some are only slightly deformed. The degree of roughening of the interface is influenced by the degree of plastic deformation. Therefore, some plastically deformed high spots of actual contact may not have a mechanical interlocking at the contacting interface strong enough to prevent relative motion within the

[3] R. Holm, *Electrical Contacts* (Almgvist and Wiksells Akademiska Handböcker, Hugo Gebers Förlag, Stockholm, 1946).
[4] F. P. Bowden and L. Leben, Proc. Roy. Soc. (London) **A169**, 391–413 (1939).
[5] Morgan, Muskat, and Reed, J. Appl. Phys. **12**, 743–52, (1941).

contacting interface. Such a pair of high spots may separate again along the original interface, contributing nothing to metal transfer or wear.

An analysis of the effect of various factors on metal transfer and wear has been made and the relation among several factors and the amount of metal transfer and wear has been formulated (by the author). A number of conclusions reached in the analysis are in good qualitative agreement with the experimental results of other investigators. These conclusions have not only applications to sliding metallic parts but also numerous applications in many special fields, e.g., metal working, metal cutting, etc. The detailed analysis will be presented in the later publication

DRY FRICTION

Dry friction between solid moving parts is a very complicated phenomenon. The frictional force is now considered to consist of several components. The mechanical component was derived from the earliest theory of friction which was suggested by Amontons and developed by Coulomb. They pictured friction as the result of geometrical interference of surface irregularities; or, more explicitly, as the tangential force required to raise hills on the upper sliding surface out of the valleys on the lower surface against the normal load. Presumably, they considered the rubbing parts as rigid bodies. It has been found that their theory is far from adequate, but the geometrical interference of surface irregularities does give rise to a usually small mechanical component of friction (also called Amontons' component or Coulomb's component), which has recently been demonstrated by Lewis and Strang.[6]

In 1929, Tomlinson[7] suggested the molecular theory of friction, according to which irreversible disturbance of molecules of the surfaces in contact is responsible for friction. His theory has not received very good experimental support. It is still uncertain whether there is a Tomlinson component of friction or not.

The "welding" theory of friction that has been developed by Holm,[8] Ernst and Merchant,[8] and Bowden[9] suggested the shearing component of friction. The essence of the "welding" theory is as follows: Continuous metallic junctions are assumed to be established between points of actual contact. The frictional force is thought to be the force required to shear off the metallic junctions.

The deformation component of friction arises from the elastic and plastic deformation of the high spots on the rubbing surfaces. When one of the rubbing surfaces slides over the other, the high spots of the surfaces will come into contact, deform, and pass over. Energy must be consumed in deforming a pair of high spots elastically. There is always an energy loss during the cycle of elastic deformation and elastic recovery. Although a great part of the elastic energy consumed in deforming a pair of high spots will be recovered during elastic recovery, yet a portion of the elastic energy will be lost due to internal friction of the material. This energy loss, though very small, also contributes a small part to the friction, and therefore yields the elastic component of friction. Energy is also consumed in deforming high spots plastically, which is almost entirely unrecoverable. This energy loss due to plastic deformation is another component of friction, the plastic deformation component. The elastic deformation component of friction has received little attention, probably because it is usually small in magnitude as compared to other components.

Among the friction components which have been cited above, the one due to shear is generally of major importance. It can easily be seen that the proposed mechanism of metal transfer and wear also explains the shearing component of friction. The difference between the "welding" theory and the proposed mechanism of metal transfer and wear is as follows: The proposed mechanism of metal transfer and wear explains the welding of the sheared-off peak to its opponent high spot as the consequence of friction, whereas in the "welding" theory, "welding" is considered as the cause of friction. One of the shortcomings of the "welding" theory which becomes immediately obvious is that it fails to explain the presence of the loose wear particles accompanying dry friction. According to the "welding" theory, continuous metallic junctions are assumed to be established between points of actual contact. If the junctions are not strong enough to prevent separation along the original interfaces, there will be neither transferred metal nor loose wear particles. If the junctions are strong enough to prevent separation along the original interfaces, there will be transferred metal, but no loose wear particles.

ADHESION AND METAL TRANSFER

According to the "welding" theory, appreciable friction and metal transfer will take place only when continuous metallic junctions can be established between two contacting surfaces. It has been suggested that the reason of formation of continuous metallic junctions is metallic adhesion. There is little question about the existence of metallic adhesion between two perfectly clean metallic surfaces which are free from any surface contamination whatsoever. However, the phenomenon of adhesion of perfectly clean metallic surfaces could not be of primary importance in friction and metal transfer, because friction does exist and metal transfer can take place in the absence of adhesion.

[6] C. R. Lewis and C. D. Strang, J. Appl. Phys. **20**, 1164–7 (1949).

[7] G. A. Tomlinson, Phil. Mag. (7) **7**, 905–39 (1929).

[8] H. Ernst and M. E. Merchant, Proc. Special Summer Conferences on Friction and Surface Finish, M.I.T., Cambridge, Massachusetts, pp. 206–11.

[9] F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids*, (Oxford University Press, London, 1950).

Therefore, as has been suggested in the beginning of this paper, adhesion is recognized as one of the secondary causes of metal transfer in this proposed new mechanism of metal transfer. The following discussion will make this argument clearer.

It is to be emphasized first that we are usually dealing with contaminated metallic surfaces because most metallic surfaces in ordinary atmosphere are always covered by a surface film which contains mainly oxides, adsorbed gases and vapors, and grease and the ordinary "cleaning" method removes only the grease. For a pair of contaminated surfaces metallic adhesion is absent. According to the calculation of Ernst and Merchant,[8] with the assumption that the frictional resistance is entirely due to shearing off metallic junctions, the static coefficient of metallic friction ($\mu = s/p$, where s is the shear stress and p is the flow pressure of the metal) is of the order of one. If metallic adhesion is the primary cause of friction, that is, the formation of metallic junctions is due to metallic adhesion, the absence of adhesion in ordinary atmosphere between two contaminated surfaces shall reduce the static friction coefficient to a negligibly small value. However, the fact is that the static coefficient for such a pair of contaminated surfaces is not negligibly small (ranging approximately from 0.2 to 1). Therefore, metallic adhesion could not be the primary cause of metal transfer and the shear component of friction.

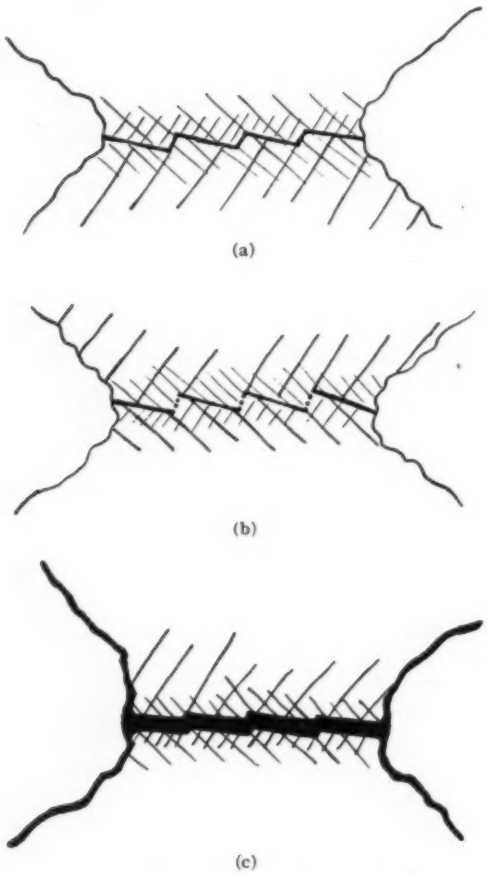


FIG. 2. Sketch showing the roughened interface of a pair of actually contacting high spots with contaminated surface. (a) For the case that the surface film is tenacious, and thin as compared to the average roughening of the interface. (b) Same as (a) except that the surface film is brittle. Establishment of metal to metal contact in the regions shown by dotted lines is possible. (c) For the case that the thickness of the surface film is close to or greater than the average depth of roughening of the interface.

Only in very special cases, we encounter the problem of rubbing two perfectly clean surfaces in vacuum. However, it is a very interesting case. If a pair of high spots on two perfectly clean surfaces of same metal come into contact in vacuum, in addition to the roughening of interface due to plastic deformation, adhesion causes continuous metallic bonds extending from one high spot into the other and the identity of the original contacting interface, though roughened, may vanish. The sheared-off peak of the high spot in this case becomes a piece of transferred blob and no loose wear particle will be formed. Experimental results have shown that the static coefficient of metallic friction of specimens cleaned with the best known laboratory technique, is much higher than one in vacuum. For example, Gwathmey, Leidheiser, and Smith[10] have reported that the static coefficient of friction under such conditions can be as high as 100 or even higher. Such experimental data have been used in supporting the idea that metallic adhesion is the basic cause of friction and metal transfer. Careful consideration reveals that the metallic junctions formed due to adhesion under a normal load in vacuum are not able to cause such extremely high friction coefficient because it is impossible to believe that the shear stresses of a metal could be as high as 100 times its flow pressure p or even more.

The real cause of extremely high friction coefficient under such conditions is the off-plane tangential force (a tangential force which is applied not in the plane of contact of the friction specimens, but at some distance away from the plane). An off-plane tangential force is equivalent to a tangential force of equal magnitude acting in the plane of contact and at the same time a moment which is equal to the product of the tangential force and its distance from the plane of contact. When an off-plane tangential force is gradually applied, the moment will cause an increase in the area of contact at the leading edge and a tendency of separation at the trailing edge. If the separation at the trailing edge is prevented by adhesion, a net increase in contacting area results. The increase in the off-plane tangential force causes an increase in the contacting area which in turn offers a greater tangential resistance. For this reason, the friction coefficient and the amount of metal transfer can become extremely high. In other words, extraordinarily large friction and metal transfer between two perfectly clean surfaces in vacuum are

[10] Gwathmey, Leidheiser, and Smith, Natl. Advisory Comm. Aeronaut. Technical Note No. 1461.

the result of the combined effect of adhesion and off-plane tangential force.

In the ordinary case of no adhesion, nothing will prevent the separation at the trailing edge, there will be no net increase in contacting area due to off-plane tangential force, and therefore, no unusually high friction coefficient and metal transfer will result.

SURFACE CONTAMINATION

Metal transfer, wear, and the shear component of friction is caused primarily by the mechanical interlocking of the roughened interface. Ordinarily, metallic surfaces are always covered with a layer of surface film which prevents metallic adhesion. The thickness and mechanical strength of the surface film has important effect on the shear resistance of the mechanically interlocked high spots. Figure 2(a) shows the roughened interface of a pair of contacting high spots for the case that the thickness of the surface film is small as compared to the average depth of roughening of the interface and, at the same time, the film is tenacious so that it remains continuous after roughening of the interface. In this case, the surface film has very little effect on the strength of the mechanically interlocked high spots. The shear component of friction is still equal to the shear stress s divided by the flow pressure p, and metal transfer is not prevented by the surface contamination provided that the diffusion process during the temperature flash is rapid enough to cause welding.

If the surface film is very brittle, it is possible that the plastic deformation causes metal to metal contact in the regions shown by dotted lines in Fig. 2(b). Metallic adhesion in these metal to metal contact regions has little effect in increasing the shear resistance of the mechanically interlocked high spots. However, as has been discussed above, the effect of metallic adhesion combined with an off-plane tangential force will cause a net increase in contacting area which, in turn, causes a higher friction as well as a larger amount of metal transfer, although adhesion itself causes little change in the shear resistance of the mechanically interlocked high spots.

If the thickness of the surface film is close to or greater than the average depth of roughening of the interface as shown in Fig. 2(c), the strength of the film has its influence on the shear resistance of the mechanically interlocked high spots. Breakage will occur inside the surface film, if the film is weaker than the strain-hardened metal near the interface. Thus, friction will be decreased, metal transfer will be greatly reduced, and wear will take place inside the surface film rather than the metallic bulk.

EXPERIMENTAL

For supporting the proposed new theory of metal transfer and wear, experiments were designed to yield direct evidence of (1) the roughening of the interface as the result of the plastic deformation, (2) the perfect matching at the roughened interface, which gives the strong mechanical interlocking effect, and (3) a definite region of severely strain-hardened metal near the interface.

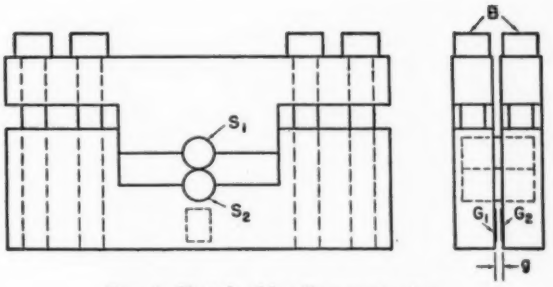


FIG. 3. Sketch of loading apparatus.

1. Apparatus

Figure 3 is a sketch of the apparatus used in the experiments. The cylindrical specimens S_1 and S_2 were seated in the semicylindrical grooves with their axes parallel. The bolts B were provided for taking over the compressive load from the testing machine, with the aid of the strain gauges G_1 and G_2, which were attached to the frame as shown. The procedure of shifting the load was as follows: The apparatus, after it had been properly assembled, was put between the anvils of the testing machine. A block of steel was put on the top of the upper frame to provide the necessary space for tightening the bolts B. Then the compressive load was applied. Suppose that the reading of the electric resistance of the gauges was i when the load applied by the testing machine had reached the desired value. The load was then decreased by five to ten percent, the electric resistance of the gauges changing correspondingly to j. The bolts were tightened with a torque wrench until the electrical resistance came back to i, indicating that the load had reached again its original value. The use of the torque wrench helped in getting a fairly uniform distribution of load. From this time on, the dial on the testing machine no longer indicated the load on the specimens. The load on the specimens was indicated by the electric resistance of the strain gauges. The load applied by the testing machine was reduced again until the electric resistance changed to j. The bolts were tightened again until the electric resistance came back to i. By repeating this procedure ten to fifteen times, the load was finally taken over from the testing machine by the bolts. Five to ten percent variation of the normal load was thus allowed in the experiments.

The apparatus was then removed from the testing machine. The specimens were kept under compression by the bolts. The specimens were cut into two in a plane perpendicular to their axes, the gap g (shown in the side view of Fig. 3) being provided for this purpose. The use of a jeweler's saw minimized the distortion.

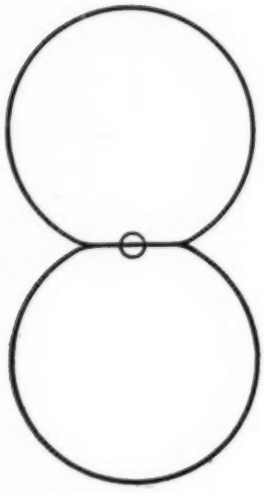


FIG. 4. Sketch showing the location on the specimen, where replica was made (indicated by the small circle).

The cross section of the specimens was then repeatedly polished and etched with a freshly prepared solution of NH_4OH plus H_2O_2 (at least four times), to remove the distorted layer due to cutting. Complete removal of this layer was determined by the x-ray back-reflection method. Thus, the intersection of the interface and the perpendicular sectioning plane could be examined, while the specimens were under compression.

2. Preparation of Specimens

The material used in the experiments was 70–30 brass in its annealed condition. The brass stock had an average grain size of 0.01 cm. Cylindrical specimens $\frac{1}{2}$ inch in diameter and 1 inch long were machined and their cylindrical surfaces were polished successively with metallographic emery papers of increasing fineness from 1/0 to 4/0. The cylindrical surfaces were then polished with 600F alundum which gave a surface with very fine scratches when examined under an optical microscope at 100× magnification.

The final polishing was carried out with 0–$\frac{1}{2}$ micron diamond dust. Multidirectional polishing technique was used. The finely polished surface had a scratch-free appearance at 100× magnification. The width and depth of scratches on the finely polished surfaces were determined in the supplementary experiment[11] as about 0.05 and 0.01 micron, respectively.

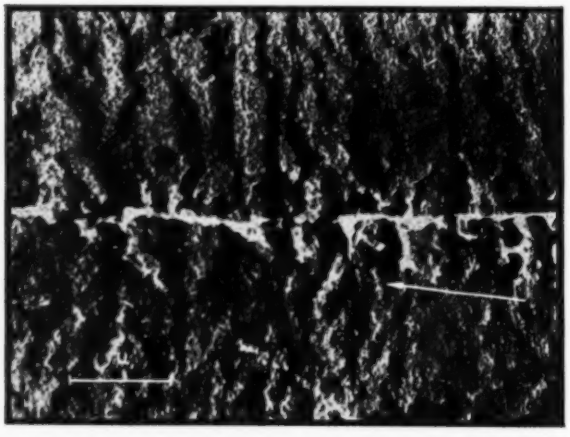


FIG. 5. Electron micrograph showing the stepwise appearance of the line of intersection of the interface and the sectioning plane. The cylindrical brass specimens were still under the compressive load of 19,000 lb. The cylindrical surfaces of the specimens of 0.01-cm grain size were finely polished with 0–1/2 micron diamond dust before being compressed. The cross section was etched with freshly prepared solution of NH_4OH plus H_2O_2. The collodion negative replica was shadowed with uranium in the direction indicated by the arrow.

RESULTS

1. Roughened Interface

A pair of annealed 70–30 brass specimens of 0.01 cm average grain size was prepared as described above and was then compressed to a total load of 19,000 pounds with a loading speed of 500 lb/min.

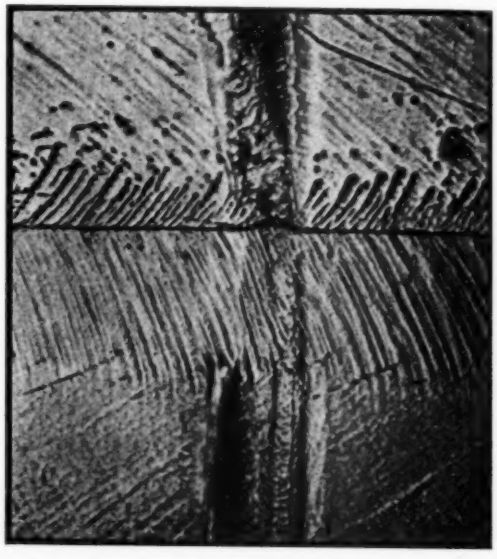


FIG. 6. Optical micrograph showing the intersection of the roughened interface and the perpendicular sectioning plane while the specimens of 0.25 cm average grain size were still kept under the maximum load of 10,000 lb reached during compression. Magnification ×1000.

The line of intersection of the interface and the perpendicular sectioning plane (use will be made hereafter of the abbreviated term, the interface line), was brought out by etching, and was examined with an optical microscope. Due to the insufficient resolving power of the optical microscope, it was not certain whether the interface was roughened or not. The interface line was then examined with the aid of the electron microscope. The collodion negative replica was made near the center of the interface line as shown in Fig. 4, and was shadowed with uranium at an angle of about arc tan $\frac{1}{3}$. Figure 5 shows the electron micrograph of the interface line. The direction of shadow-casting is indicated by the arrow. The roughening of

[11] I. M. Feng, J. Appl. Phys. **22**, 820 (1951).

the interface due to plastic deformation is shown in Fig. 5 by the stepwise appearance of the interface line. Figure 5 also shows the intimate contact and the perfect matching of the upper and the lower specimens at the roughened interface.

As has already been determined in previous experiments,[11] the scratches on the finely polished cylindrical surfaces had only a width of 0.05 micron and a depth of 0.01 micron, and would not give a sharp stepwise appearance.

2. Roughened Interface of Specimens of Very Large Grain Size

The same 70–30 brass stock was annealed at 1550°F for 8 hr to give a very large average grain size of 0.25 cm. The same experiment was repeated with a compressive load of 10,000 lb.

The optical micrograph, Fig. 6, shows the zig-zag shape of the interface line. The very large grain brought the roughening into the resolving power of the optical microscope. This picture also shows that there is a definite region of most severely strain-hardened metal near the interface as indicated by the strain lines, and this strain-hardened region had different depths in different grains due to their different crystallographic orientation and their different conditions of constraint imposed by their neighboring grains.

The vertical mark in the middle of the picture was made by a fine, sharp needle. It is interesting to notice that the scratch is narrower inside the most severely strain-hardened region than outside that particular region. The width of the scratch inside the most severely strain-hardened region decreases as the scratch approaches the interface. It serves to indicate the variation of the degree of strain-hardening.

The interface line was also examined with the aid of the electron microscope. The collodion negative replicas were made and shadowed with uranium at an angle of about arc tan $\frac{1}{3}$. Figures 7 and 8 are the electron micrographs of the interface line. They show two different degrees of roughening at two different locations inside two different grains. They demonstrate the effect of the crystallographic orientation of the grains and the constraint imposed by the neighboring grains.

3. The Effect of Removing the Load on the Roughened Interface

In the foregoing experiments, the investigation was limited to the line of intersection of the interface and the perpendicular sectioning plane. In order to examine the entire interface, the compressive load on the specimens must first be removed. The question then arises as to whether there is any change in the roughening of the interface during the removal of the compressive load. The specimens of very large grain size that were used in the previous experiment were put back on the testing machine and the compressive load was shifted back from the bolts to the machine. The compressive load was then removed at a rate of 500 lb/min. Figure 9 is an optical micrograph showing the appearance of the roughened interfaces after the removal of the compressive load.

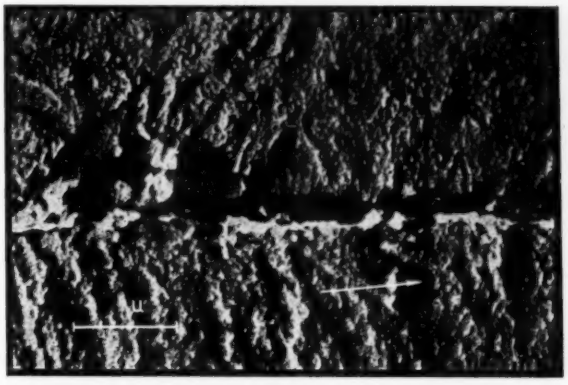


FIG. 7. Same as Fig. 5 except that the grain size was 0.25 cm and the compressive load was 10,000 lb. The arrow indicates the direction of shadow casting.

Figures 6 and 9 illustrate the identical area on the cross section of the specimens, before and after the removal of the compressive load. They show no appreciable change in the roughening of the interface. Because of the limited resolving power of the optical microscope, this experimental result does not mean that there was absolutely no change in the roughening of the interface during the removal of the compressive load.

4. Appearance of the Roughened Interface After Separation

By following the same procedure as that used in removing the compressive load on the specimens of very large grain size (0.25-cm average size), the compressive load on the specimens of 0.01-cm average grain size was removed and the specimens were then separated. Figures 10 and 11(a) are two representative optical

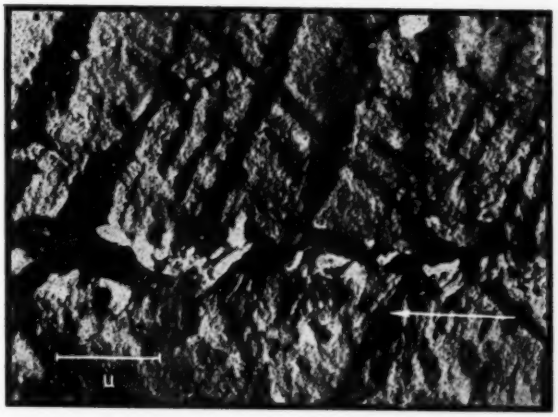


FIG. 8. Same as Fig. 5 except that the grain size was 0.25 cm and the compressive load was 10,000 lb. The arrow indicates the direction of shadow casting.

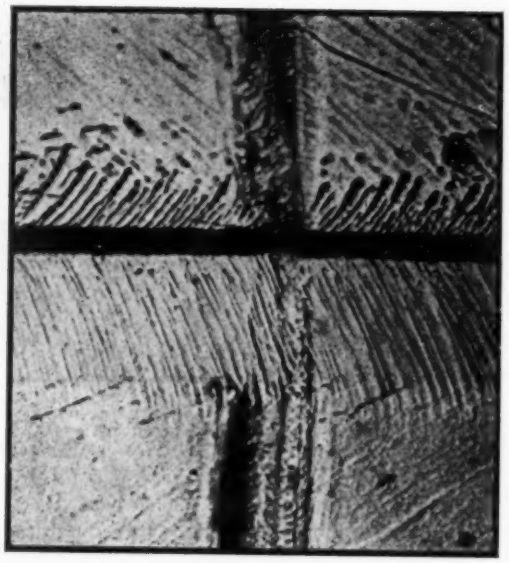


FIG. 9. Optical micrograph showing the appearance of the intersection of the interfaces and the perpendicular sectioning plane after the removal of the compressive load. Compare this picture with Fig. 6. They illustrate the identical area on the cross section of the specimens, before and after the removal of the load. Magnification ×1000.

FIG. 10. Optical micrograph showing the complex appearance of the roughened interface of the cylindrical brass specimen after the removal of the compressive load (19,000 lb). The cylindrical surfaces of the specimens of 0.01-cm grain size were finely polished with 0–1/2 micron diamond dust before being compressed together. Note the intersection of the twinning bands (middle left), the crossing of the twinning bands and the grain boundaries and the passing of the slip lines through the twinning bands (central), which were the result of simultaneous plastic deformation of two contacting surfaces. Magnification ×750.

FIG. 11(a). Optical micrograph showing the complex appearance of another part of the roughened interface of the cylindrical brass specimen after the removal of the compressive load (19,000 lb). The cylindrical surfaces of the specimens of 0.01-cm grain size were finely polished with 0–1/2 micron diamond dust before being compressed together. Note the passing of the grain boundaries through grains, and the crossing of slip lines and grain boundaries, which were the result of simultaneous plastic deformation of two contacting surfaces. Magnification ×750.

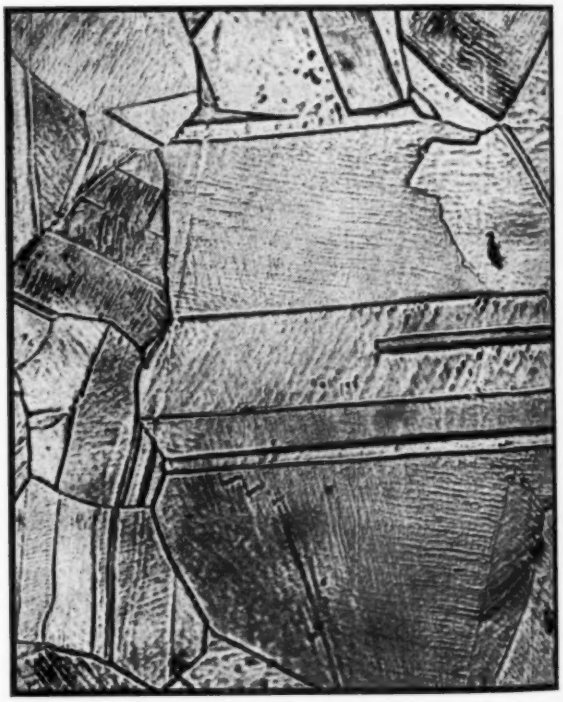


FIG. 11(b). Optical micrograph of the same area on the specimen as that of Fig. 11(a), after etching with NH_4OH plus H_2O_2. Note the disappearance of the impression. Magnification ×750.

pictures of the roughened interface. The grain boundaries and the twinning boundaries were brought out by plastic deformation. Careful observation reveals the following details:

(1) Twinning bands appear to cross one another (middle left of Fig. 10).
(2) Twinning bands appear to lie across grain boundaries (upper left corner of Fig. 10).
(3) Grain boundaries appear to pass through grains [Fig. 11(a)].
(4) Slip lines appear to run across grain boundaries [center of Fig. 11(a)].
(5) Slip lines appear not to stop at twinning planes (the central part of Fig. 10).

The complex appearance of the roughened interface indicates that the roughening of one surface due to plastic deformation made its impression on the opposite contacting surface.

Brass is of the face-centered cubic structure which has four sets of slip planes of the (111) type. Only two sets of mutually perpendicular slip lines can appear on any (100) plane. In the upper middle part of Fig. 10, there are more than two sets of slip lines among which two sets are mutually perpendicular. This clearly indicates that they can not all belong to the same specimen. Only three sets of slip lines forming equilateral triangles can show on planes of the (111) type. In the lower middle of Fig. 10, there are three sets of slip lines forming equilateral triangles, with an additional set of slip lines. It is obvious that either these three sets of slip lines forming equilateral triangles or the additional set must be the impression from another surface. The analysis of the angles of intersection of the slip lines leads to the same conclusion that the roughening of one surface due to plastic deformation made its impression on the opposite contacting surface and produced a complex appearance of the roughened interface.

The interface was then etched with freshly prepared solution of NH_4OH plus H_2O_2. Figure 11(b) shows the same area on the specimen after etching as that of Fig. 11(a). The disappearance of the impressed grain boundaries and slip pattern confirms the above conclusion.

Figure 12 was taken from the specimens of 0.25-cm average grain size. It shows the areas on the upper and lower surfaces that were in contact during plastic deformation. It helps in visualizing directly the perfect matching of the roughened interface, which gives a strong mechanical interlocking effect.

FIG. 12. Optical micrograph showing the areas on the upper and lower surfaces that were in contact during plastic deformation. The total compressive load on the specimens of 0.25-cm grain size reached during compression was 10,000 lb. This picture helps in visualizing directly the perfect matching of the roughened interface, which gives a strong mechanical interlocking effect. Magnification ×100.

ACKNOWLEDGMENTS

Particular thanks are due to Professor E. Katz of the University of Michigan for his great interest, stimulating discussion, and constant encouragement without which this work could not have been completed. The author also wishes to thank Professors R. C. Porter, E. T. Vincent, Clair Upthegrove, Lars Thomassen, and G. V. Edmonson of the University of Michigan, and Professor R. C. Williams of the University of California for their advice and help. The technical assistance given by Mr. R. N. Clark should also be acknowledged. The author must express his gratitude to the Horace H. Rackham School of Graduate Studies of the University of Michigan for the Horace H. Rackham Predoctoral Fellowship.

A Treatment of the Viscosity of Concentrated Suspensions*

ROBERT SIMHA
Department of Chemical Engineering, New York University, University Heights, New York
(Received April 17, 1952)

In current theories of the concentration dependence the solute molecules are treated as point centers of disturbance. The finite size that produces a shielding of two interacting particles in the presence of intervening ones reduces the interaction calculated in the point approximation and thus reduces the discrepancy between experiment and theory. An approximate treatment for a spherical solute based on the cage model, and hence more adequate at high concentrations, is presented. It operates with a spherical enclosure around a central particle. The position of this shell reflects the relative extent to which particles beyond nearest neighbors can hydrodynamically interact with the central molecule. Thus, a parameter f appears in the theory which increases slowly with concentration and approaches a limit corresponding to the coincidence of the shell with the wall of the nearest neighbor cage at high concentrations. In the framework of the present theory, f must be regarded as a semiempirical quantity. The hydrodynamic problem can be handled rigorously. The relative viscosity calculated is in satisfactory agreement over a range of concentrations with empirical expressions successfully applied to pertinent systems. The initial value of f derived from the observed coefficient of the quadratic term is also reasonable.

IT has been found, in general, that the calculated or anticipated concentration dependence of the viscosity in relatively dilute macromolecular solutions exceeds the observed effect. This holds for flexible molecules as well as compact particles. In the case of polymer solutions, reasons have been advanced for this discrepancy.[1] The principal one is the mutual compression of the average coils in a good solvent at rather low concentrations, which reduces the effect of hydrodynamic interaction. In the following we deal exclusively with spherical suspensions, which are most suitable for the study of hydrodynamic factors. However, the results allow the drawing of some conclusions about other solutes also.

For this case, considerations based on earlier work[2] led to a term $14.1\varphi^2$ to be added to Einstein's formula for the relative viscosity.[3] φ is the volume fraction of the solute.[4,5] A consideration of the mutual volume exclusion[6,7] at finite concentrations reduces the value to 12.6. Experiments on various systems, including glass spheres, on the other hand, indicate values between 7 and 10. Neither aggregation nor compression effects considered for polymer solutions can be invoked here. The former would lead to an increase of the quadratic term, since the doublet has a higher intrinsic viscosity than the singlet.

There is, however, a factor that is particularly important for compact particles and has been omitted from consideration in all theories so far advanced. Hitherto each particle has been treated as a "force" center with a (long-range) "potential" which in the viscosity problem decays as a^3/r^3 and in sedimentation as a/r, where a is the particle radius. This is adequate as long as the molecules can be regarded as points. Their finite size, however, introduces a shielding effect which becomes increasingly important with increasing concentration and reduces the result calculated in the point approximation. The effect can be characterized by a length L analogous to the shielding length introduced in familiar theories of intrinsic viscosity. L is roughly the distance over which the hydrodynamic disturbance produced by one molecule can affect a second one in the presence of intervening particles. The critical ratio is R_{12}/L, where R_{12} represents the average intermolecular distance. The ratio must vanish at high dilution and become of the order of unity at the other limit. By analogy with the Debye-Bueche theory of intrinsic viscosity,[8] one deduces a proportionality of this ratio with $\varphi^{\frac{1}{2}}$. In view of the low density of a polymer coil, it is possible that this effect would not alter significantly the magnitude of the intrinsic viscosity in this case.

A rigorous treatment of this screening effect, even for dilute solutions, represents no easy problem. We shall

* Presented at the Meeting of the American Physical Society, Division of High-Polymer Physics, March 20–22, 1952, Columbus, Ohio.

[1] Weissberg, Simha, and Rothman, J. Research Natl. Bur. Standards **47**, 298 (1951).

[2] E. Guth and R. Simha, Kolloid-Z. **74**, 266 (1936).

[3] O. Gold, thesis, Vienna (1937).

[4] The discrepancy between the value 109/14 derived in reference 2 and 14.1 is caused by the omission of some terms arising from u_2, v_2, w_2 on page 273 in the equation for $u_i^{(2)}$ at the bottom of page 274.

[5] There is a question about the manner in which the requisite summations of the hydrodynamic disturbance over the volume of the solution have been carried out. It arises already in Einstein's published derivation and has to do with the choice of the domain of integration which includes a singularity. The surface integral and the effective components of dilatation of the solution finally evaluated lead, of course, to the correct result. This can also be obtained from a volume integral (which may or may not be transformed into a surface integral) by the proper choice of domain. In the case of nonspherical particles, it has been shown by J. Riseman and R. Ullman [J. Chem. Phys. **19**, 578 (1951)] and R. Simha [J. Research Natl. Bur. Standards **42**, 409 (1949)] that the surface integral over a sphere concentric with the point for which the total disturbance is computed yields practically the same result for the quadratic term as Burgers' method of summation.

[6] J. M. Burgers (unpublished work).

[7] N. Saito, J. Phys. Soc. (Japan) **5**, 4 (1950).

[8] P. Debye and A. M. Bueche, J. Chem. Phys. **16**, 573 (1948).

adopt a simplified model, derived from the familiar cage model of liquids and solutions. At high concentrations, only the interaction between the trapped particle and its nearest neighbors is important. If the cage is replaced by a concentric spherical enclosure, the hydrodynamic problem can be solved rigorously. The approximations reside in the deficiencies of the model at low concentrations. Actually a similar model has been used by Cunningham[9] in his study of the modifications of Stokes' law required in connection with the evaluation of the elementary electronic charge. In the case of sedimentation, the ratio between effective shell radius b and R_{12} must vary more with increasing concentration than in viscosity, because of the $1/r$-relation in the former and because the model is less adequate.

As will be shown, the model serves to illustrate the sensitivity of the quadratic concentration term toward the screening effect in addition to yielding an expression for the viscosity of highly concentrated spherical suspensions.

MATHEMATICAL DEVELOPMENT

In the absence of suspended particles, let the solvent be subject to a simple shearing motion with a velocity gradient q in the x_2-x_3 plane:

$$u_i=\delta_{i3}qx_2. \tag{1}$$

Since only spherical particles are considered, the rotational part of the flow (1) is not disturbed and only the dilatational contribution

$$u_i^{(0)}=q/2(\delta_{i3}x_2+\delta_{i2}x_3) \tag{1a}$$

need to be taken into account.

Einstein's solution, to be added to (1a) satisfies the equations of motion for the velocity components u_i and the pressure p, the continuity equation for an incompressible fluid

$$\nabla^2u_i=\partial p/\partial x_i;\quad \partial u_k/\partial x_k=0, \tag{2}$$

and the boundary conditions

$$u_i=\begin{cases}-u_i^{(0)}, & r=a\\ 0, & r=\infty\end{cases}\qquad r^2=x_k\cdot x_k.$$

The appropriate solution is:

$$\begin{aligned}u_i^{(1)}&=-5/2qa^3x_ix_2x_3/r^5(1-a^2/r^2)\\ &\quad -q/2a^5/r^5(\delta_{i3}x_2+\delta_{i2}x_3),\\ p&=-5\eta_0qa^3/r^5x_2x_3+\text{ const.}\end{aligned} \tag{3}$$

η_0 represents the viscosity of the solvent. Equation (3) or analogous relations form the basis for all theories of the concentration effect. We now place around the particle $r=a$ a concentric sphere $r=b$ which is to replace in its effect on a the interaction of all outlying portions of the suspension with the central molecule. The velocity distribution (3) will then be modified in such a manner that the resulting field $u_i^{(2)}$ fulfills the conditions:

$$u_i^{(2)}=\begin{cases}u_i^{(1)}=-u_i^{(0)}, & r=a\\ 0, & r=b\end{cases} \tag{4}$$

As shown in Sec. 1 of the Appendix, $u_i^{(2)}$ takes the form

$$\begin{aligned}u_i^{(2)}&=-5/2qa^3\cdot\lambda[1-a^2/r^2\cdot b^2(b^5-a^5)/(b^7-a^7)]\\ &\quad\times x_ix_2x_3/r^5+5/2qa^3\lambda(b^2-a^2)/(b^7-a^7)x_ix_2x_3\\ &\quad-q/2a^5/r^5\cdot\lambda b^2(b^5-a^5)/(b^7-a^7)(\delta_{i3}x_2+\delta_{i2}x_3)\\ &\quad-25/8qa^3\lambda/(b^7-a^7)[(b^2-a^2)r^2\\ &\quad-(25b^7-21b^5a^2-4a^7)/25b^3](\delta_{i3}x_2+\delta_{i2}x_3)\\ p&=-5\eta_0qa^3/r^5\cdot\lambda\cdot x_2x_3-105/4\eta_0qa^3\cdot\lambda\cdot\\ &\quad(b^2-a^2)/(b^7-a^7)x_2x_3+\text{ const.}\end{aligned} \tag{5}$$

λ is a function of the ratio $y=a/b$.

$$\lambda(y)=4(1-y^7)/[4(1+y^{10})-25y^3(1+y^4)+42y^5]. \tag{6}$$

It turns out that the viscosity behavior is determined by the function λ. Hence, it is desirable to consider the following expansions:

$$\left.\begin{aligned}\lambda&\doteq 1+25/4y^3-21/2y^5+625/16y^6\\ &\quad+21/4y^7-525/4y^8+0(y^9);y\to 0\\ \lambda&\doteq 4/25.1/\delta^3[1+3\delta+21/5\cdot\delta^2+17/5\cdot\delta^3\\ &\quad+0(\delta^4)],\delta=(1-y)/y;y\to 1.\end{aligned}\right\} \tag{6a}$$

The second expansion is the more useful one and gives sufficiently correct results down to values of $y=\frac{1}{2}$. For $b\to\infty$, Eqs. (5) and (3) become identical, as they must.

Next the energy dW/dt, dissipated per unit time within a cell by virtue of the flow disturbance (5), must be computed. This can be done by integrating the dissipation function over the cell volume (see Sec. 2, Appendix), or by means of the surface integral

$$dW/dt=\int p_{ri}\cdot u_i\cdot dS=\int p_{ki}\cdot u_i\cdot x_k/r\cdot dS,$$

where p_{ki} represents the viscous stress tensor and p_{ri} the radial stresses. The integration is to be extended over the outer surface only, where u_i reduces to $u_i^{(0)}$, since the contribution from the boundary $r=a$ vanishes. Utilizing the expressions given in Sec. 1 of the Appendix, one finds for the contribution of a cell

$$dW/dt=4\pi/3b^3q\left[\frac{b^2}{5}K+2\eta_0k-(3F/10b^3)\right].$$

By substitution of the values for the constants and inclusion of the dissipation of the pure solvent we obtain the following equation for the relative viscosity of a solution of N spheres in a volume V:

$$\eta/\eta_0=1+5/2\lambda\varphi, \tag{7}$$

where $\varphi=4\pi/3a^3N/V$ represents the volume fraction occupied by the suspended particles.

[9] E. Cunningham, Proc. Roy. Soc. (London) **83A**, 357 (1910).

Thus the quantity λ measures the interaction between a given particle and all other ones. Figure 1 shows λ as a function of the ratio y. As soon as the radius of the cell becomes smaller than the diameter of a particle, the viscosity starts to rise sharply. As anticipated, this increase is more moderate than that in sedimentation. There the expansion of the interaction factor starts with the first power of y, reflecting the difference between translation and shear behavior. Only in the limit $y\rightarrow 1$ do the two factors become proportional to each other, and the apparent viscosity for the sedimentation process tends toward a value equal to 10/3 times the actual viscosity of the suspension.

CONCENTRATION DEPENDENCE

The position of the cell wall $r=b$ will vary with the concentration. At high dilution, where the screening effect does not prevent interactions reaching considerably beyond nearest neighbors, the shell that is to replace the interaction, must necessarily be located at $b<R_{12}$. As the particles come closer together, shielding increases, and b/R_{12} slowly increases and tends toward a limit. As shown below, the experimental data at high dilution indicate that this anticipated shift involves only a relatively small fraction of the mutual distance R_{12}. This is not surprising in view of the fact that the interaction varies rapidly enough, namely, as $1/r^3$.

We can quite generally assume either

$$a^3/b^3=y^3=\varphi/f^3 \tag{8a}$$

or

$$y^3=\varphi/[f^3(1-\varphi^{\frac{1}{3}}/f)^3]. \tag{8b}$$

Equation (8a) corresponds to the assumption that b is proportional to R_{12}, while (8b) implies proportionality with $(R_{12}-a)$. At low concentrations, the difference between the two alternatives is small. At high concentrations, (8b) seems more adequate. According to (8b), the maximum concentration φ_{max}, corresponding to close packing and $y=1$, is given by

$$\varphi_{max}=f^3/8.$$

Thus, for an hexagonal or simple cubic arrangement, $f=1.81$ or 1.61, respectively, if the shell b is identified with the wall of the cage formed by the nearest neighbors, $b=R_{12}-a$. At low φ, therefore, f will be smaller than the above values.

Substitution of (8) into (6a) and (7) gives for low concentrations the expansions

$$\eta_{sp}/\varphi\doteq 5/2\{1+\frac{25}{4f^3}\varphi-21/2f^5\varphi^{5/3}+625/16f^6\varphi^2+0(\varphi^{7/3}/f^7)\} \tag{7a}$$

$$\eta_{sp}/\varphi\doteq 5/2\{1+25/4f^3\varphi+75/4f^4\varphi^{4/3}+27/f^5\varphi^{5/3}+785/16f^6\varphi^2+0(\varphi^{7/3}/f^7) \tag{7b}$$

and

$$\lim_{\varphi\rightarrow\varphi_{max}}\eta/\eta_0=(54/5f^3)[\varphi^2/(1-\varphi/\varphi_{max})^3]. \tag{7c}$$

DISCUSSION

Equations (7a) and (7b) show the pronounced influence of the screening effect on the initial slope of the viscosity-concentration curve. The coefficient of the linear term is $15.6/f^3$. Regardless of the specific arrangement assumed, random, hexagonal or any other, the coefficient remains small, of the order of 2–4, if interactions with other than nearest neighbors are excluded. This, of course, is too stringent a condition to impose. An observed coefficient of 7–8 corresponds to $f\doteq 1.31$–1.25. Now $f=1$ means $b\doteq R_{12}/2$, one of the assumptions made by Cunningham,[9] and $f=1.31$–1.25 gives $b\approx 0.72R_{12}-0.69R_{12}$, which is reasonable as a measure of the extent of interaction. Thus, over the whole range of φ, f will change from a value around 1.3 to less than 2. While the theory explains the discrepancy between previously calculated and experimentally found initial slopes, the higher terms in the expansions (7a) and (7b) should not be taken too literally in view of the deficiency of the model at low concentrations. Nevertheless, the appearance of a quasicrystalline structure in the solution and the resulting appreciable shielding will give rise to the appearance of fractional powers of φ and presumably negative coefficients. This is reminiscent of the situation in ionic solutions and is fundamentally due to the same reason, the long range character of the interaction. Incidentally, the disturbance of the relative arrangement of the solute molecules by the applied shear represents another source for non-Newtonian flow in concentrated systems, even when no tearing up of actual aggregates is involved.

Equation (7c) may be compared with Eilers'[10] empirical expression which reduces for high relative viscosity to

$$\eta_r=(25/16)[\varphi^2/(1-\varphi/\varphi_{max})^2].$$

Mooney[11] derived and applied successfully to exper-

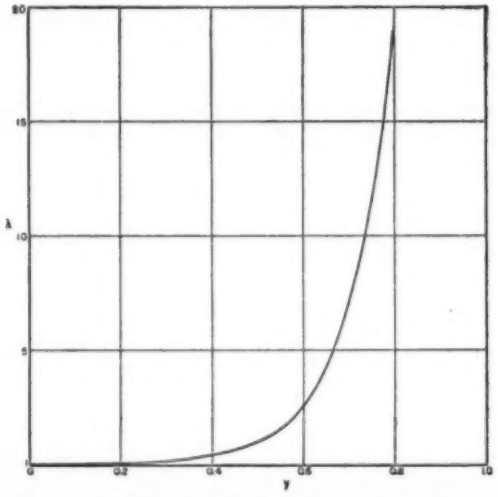


FIG. 1. Interaction factor λ as a function of the ratio $y=a/b$.

[9] H. Eilers, Kolloid-Z. **102**, 154 (1943).
[11] M. Mooney, J. Colloid Sci. **6**, 162 (1951).

imental data on various suspensions the equation

$$\eta_r = \exp[2.5\varphi/(1-k\varphi)].$$

In the process of fitting data it turned out that the constant k assumes values determined principally by the geometric packing of the spheres. This led Mooney to the conclusion[12] that hydrodynamic interaction would give an average effect that is nearly equivalent to a geometric crowding factor. The present considerations indicate the importance of the particle arrangement in influencing the magnitude of the hydrodynamic interaction which, however, is the primary effect.[13]

Figure 2 shows a comparison of our theory, Eqs. (7) and (8b) with Mooney's and Eilers' (complete) expressions, assuming $f=1.81$ or $k=1.35$. This gives for the coefficient of the quadratic term in η/η_0, respectively, 2.64, 6.4, and 4.94. Nevertheless, the reduced specific viscosities do not differ much in the range of 2–9 percent because of the contribution of the fractional powers in (7b). The limiting values for $\varphi \to \varphi_{max}$ are, of course, very different. In the intermediate range our equation is close to Mooney's formula, which, as mentioned before, reproduces the experimental data well. We must recall again, that k is empirically determined. The value of f need not be expected to correspond to this k. It would be smaller first, raising the curve in Fig. 2, and gradually approach the value used in the graph. Not much of a change in f is needed, even at high concentrations, to make Mooney's and our values identical. For instance, at $\varphi=40$ percent in Fig. 2, a change from 1.81 to 1.71 in f suffices to produce the necessary shift in the concentration scale of our expression to 38 percent. A study of the behavior of the Dow polystyrene latices will be of interest in connection with this theory.

An extension of the theory to polydisperse systems requires additional simplifications. An obvious assumption is to neglect fluctuations, assume that the local composition of the cell is identical with the over-all composition of the mixture, and introduce an average value $\langle b(\alpha)\rangle$ for a given central particle of species α. However, since the factor f also depends on the size distribution, we have arrived at no definite expression for the influence of polydispersity on the initial slope. It seems quite clear that for any type of solute, the screening effect leads to a dependence of the quadratic and higher concentration terms in the expression for η_{sp} on the size distribution. However, the effect may not be large. Recent considerations[14] indicate that the addition of large particles to a solution of small ones results in a term in η_{sp} equal to $35/8\varphi^2$, i.e., smaller than accepted here for a homogenous system.

[12] M. Mooney (private communication).

[13] The exponential relation is obtained as the solution of the functional equation $\eta_r(\varphi_1/1-k\varphi_2)\times\eta_r(\varphi_2/1-k\varphi_1)=\eta_r(\varphi_1+\varphi_2)$, where k is assumed *a priori* to absorb both volume exclusion and hydrodynamic interaction. It is not obvious that this can be represented in just that form, particularly at intermediate φ's. Furthermore, the interaction between components 1 and 2 should affect the validity of the product formula.

[14] H. C. Brinkman, J. Chem. Phys. **20**, 571 (1952).

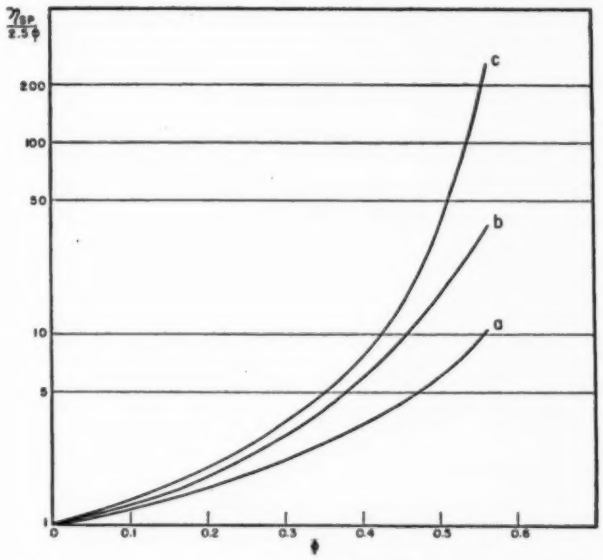


FIG. 2. Comparison of reduced specific viscosity, η_{sp}/φ, calculated from Eilers' (see reference 10) (a) and Mooney's (see reference 11) (c) expressions, with Eqs. (7) and (8b), curve b, assuming $\varphi_{max}=0.74$.

One might consider a direct application of these results to solutions of flexible macromolecules with proper allowance for the compression effect, as φ approaches φ_{max}; that is, the parameter f experiences an additional variation with concentration which depends on the thermodynamic conditions in the solution.[1] Although the model of a compact sphere with a volume equal to the encompassed volume of the coil is rather adequate for a discussion of the intrinsic viscosity, the detailed application to higher concentrations, involved here, is of somewhat questionable significance. In the analysis of polystyrene solutions,[1] it was assumed that for a fixed coil volume, the relative viscosity can be expressed by Arrhenius' relation

$$\eta_r = e^{[\eta]c}.$$

The introduction of an expression of type (7c) instead of the above would improve the numerical agreement and change the value of the extension factor α deduced from the observed concentration dependence in the right direction.

It is fair to say in conclusion that the quantitative interpretation of the concentration dependence of viscosity has reached the same level as that of the thermodynamic activity of macromolecular solutions. In both cases the description involves semi-empirical parameters. In respect to highly concentrated polymer solutions, however, the statistical mechanical theory is more advanced. There, hydrodynamic methods loose their applicability, and we still require an adequate theory of melt viscosity as a starting point.

APPENDIX

1. *Derivation of Eq. (5) fulfilling boundary condition (4).*

On the basis of Lamb's general solution of (2) in terms of spherical harmonics,[15] we choose

$$p_{-3}=Fx_2x_3/r^5, \quad p_2=Kx_2x_3,$$
$$\Phi_{-3}=fx_2x_3/r^5. \quad \Phi_2=kx_2x_3.$$

This leads to

$$\eta_0 u_i=(F/2-5f\eta_0/r^2)x_ix_2x_3/r^5-2/21Kx_ix_2x_3 + (f\eta_0/r^5+5/42Kr^2+k\eta_0)(\delta_{i3}x_2+\delta_{i2}x_3).$$

From the boundary conditions one deduces

$$F=-5qa^3\eta_0\lambda, \quad f=-q/2a^5b^2(b^5-a^5)/(b^7-a^7)\cdot\lambda,$$
$$K=-105/4\eta_0qa^3[(b^2-a^2)/(b^7-a^7)]\lambda,$$
$$k=q/8a^3/b^3\frac{(25b^7-21b^5a^2-4a^7)}{(b^7-a^7)}\lambda.$$

This leads immediately to Eqs. (5) and (6).

[15] H. Lamb, *Hydrodynamics* (Dover Publications, New York, 1945), 6th edition, pp. 596 and 598.

2. *Identity of surface integral dW/dt given in text with volume integral over Rayleigh's dissipation function.*

According to Gauss' theorem

$$\int p_{ki}\cdot u_i\cdot x_k/r\cdot dS=\int(\partial/\partial x_k)(p_{ki}\cdot u_i)d\tau.$$

But:

$$(\partial/\partial x_k)(p_{ki}\cdot u_i)=u_i\partial p_{ki}/\partial x_k+p_{ki}\partial u_i/\partial x_k.$$

In the absence of external forces the tensor divergence vanishes for the steady, inertialess motion considered here. Furthermore,

$$p_{ki}\cdot\partial u_i/\partial x_k=-p\delta_{ik}\partial u_i/\partial x_k + \eta_0(\partial u_i/\partial x_k+\partial u_k/\partial x_i)\partial u_i/\partial x_k.$$

This reduces to the last term and proves the identity.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

The Effect of Accelerating Voltage and Specimen Morphology on Electron Diffraction Patterns*

S. G. Ellis
Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey
(Received May 9, 1952)

Using empirical relations established in the study of electron diffraction patterns of uniform thin films of aluminum, the contrast between the most intense ring and the background is calculated for other morphologies as a function of the accelerating voltage on the camera.

In all cases the contrast increases with voltage. The gain is marked if the specimen is nowhere very thick. For specimens that have very thick regions the contrast increases approximately as the square root of the voltage in the range 50 to 150 kv. The situation in reflection-diffraction in general lies between these extremes. Increasing the voltage has a more limited effect on increasing the range of specimen thicknesses that can be examined than had been previously supposed.

INTRODUCTION

IT has been known for a long time that the contrast in the electron diffraction pattern from a polycrystalline thin film increases with electron speed and with a reduction in film thickness. "Contrast" may be defined with the aid of Fig. 1,

$$C=I_R/I_B, \quad (1)$$

where I_R is the current entering a given ring of the pattern and I_B is the current entering the diffuse background at the ring. In practice it is easier to measure the peak contrast C', where

$$C'=I_R'/I_B'; \quad (2)$$

I_R' is the peak current density in the ring, and I_B' is the background current density at the ring.

The form of the variation of contrast with accelerating voltage and specimen thickness is of obvious importance to the designer of an electron diffraction camera since it will enable him to judge the desirability of using high accelerating voltages. The same information may be valuable to the users of electron diffraction cameras since it will indicate the limits of electron diffraction technique.

Möllenstedt[1] has measured as a function of voltage the greatest thickness of aluminum film that will give diffraction rings visible on a fluorescent screen. When one ring could be seen on the screen, several could be seen on a photographic plate that had been directly exposed to the electron beam. This maximum thickness of aluminum that could be examined increased less rapidly than the accelerating voltage (perhaps as $V^{\frac{1}{2}}$) for voltages up to 150 kv; beyond this it increased more rapidly than the voltage (perhaps as V^2). Möllenstedt

* Parts of this paper were presented at the National Bureau of Standards Symposium on Electron Physics, Washington, November 5–7, 1951, and at the Philadelphia meeting of the Electron Microscope Society of America, November 8–10, 1951.

[1] G. Möllenstedt, Nachr. Wiss. Gottingen (No. 1) 83 (1946).

was able to obtain patterns from an aluminum film 2 microns thick at 600 kv.

Interesting though they are, Möllenstedt's results are not suited to generalization since the characteristics of the fluorescent screen and photographic plates used are not recorded. It was therefore decided to measure the peak contrast C' as a function of film thickness and accelerating voltage. This work is being reported elsewhere.[2] In the course of the work it was found that by studying the variation of I_R' and I_B' separately with film thickness some insight could be gained into the nature of the plural electron scattering which ultimately limits the contrast. In particular, empirical relations were found for I_R' and I_B':

$$I_R' = I_0' k_1' e^{-Kt} \cdot t, \tag{3}$$

$$I_B' = I_0' k_2' e^{-Kt}(t + \alpha t^n). \tag{4}$$

In these equations I_0' = current density in the illuminating beam, t = film thickness, and k_1', k_2', K, α, and n are parameters that depend on the material of the diffracting film, its degree of orientation, and the accelerating voltage. It was possible to obtain numerical values for k_1'/k_2', K, and α. The nearest integer to the observed value of n was taken in order to simplify the calculations which follow.

Equations (3) and (4) make it possible to calculate the peak contrast C' for specimen morphologies other than the uniform thin film of thickness t, used in the experimental determination of the parameters. The purpose of this paper is to outline the calculations and give the results that have been obtained.

Before proceeding to the calculations it may be worthwhile to review some of the qualitative changes that occur in the diffraction patterns of thin polycrystalline films on changing the film thickness or the accelerating voltage.

For a fixed accelerating voltage V, an increase in film thickness t results in the following:

(a) A less rapid variation in the background current density. If R is the radius at which I_B' is measured, $|dI_B'/dR|$ becomes less on increasing t, at a given value of R.

(b) The ring intensity is reduced. I_R' becomes smaller, but I_B' remains approximately constant over the range in which rings can be observed so that C' is reduced. This point is considered in more detail below.

(c) The ring width remains unchanged within the limits imposed by the necessarily large errors of measurement when the contrast is low.

If t is constant and V is increased:

(d) $|dI_B'/dR|$ is increased, at given value of $\sin\theta/\lambda$.

(e) There is a well marked increase in contrast C'.

(f) The ring width is reduced for those rings in which some of the broadening is due to refractive effects. The reduction is both absolute and relative in the sense that for given $\sin\theta/\lambda$, R (the ring radius) and $\Delta R/R$ are reduced (here ΔR is the line width). The rings become smaller and sharper.

[2] S. G. Ellis, Phys. Rev. **87**, 972 (1952).

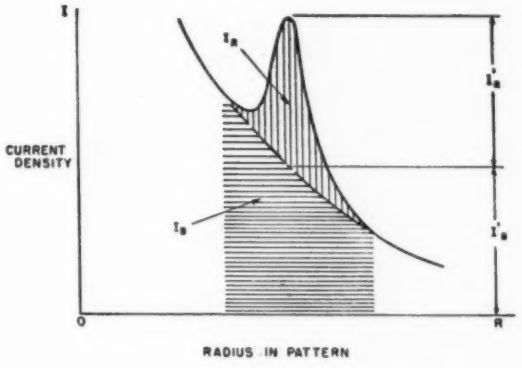


FIG. 1.

THE INFLUENCE OF MORPHOLOGY

It is very difficult to assess the value of using higher accelerating voltages on electron diffraction cameras by making direct experimental comparisons at different voltages. This is because the morphology of the specimen has a pronounced effect on the contrast, and the morphology is not generally under control. Alternatively, we can use Eqs. (3) and (4) to calculate the contrast that would be obtained with specimens of simple morphology. When this is done certain generalizations become apparent.

The term morphology as used below denotes not the crystal shape but the external form of a polycrystalline body. It is supposed that the individual crystals are small and are randomly oriented. Where numerical results are quoted it must be supposed that the degree of orientation is that which might be found in the type of evaporated thin film used in the experimental determination of the scattering parameters.

If ds is the element of cross-sectional area of the electron beam where the specimen thickness is t, and if β is the cross-sectional area of that part of the illuminating beam which strikes the specimen, then

$$I_R' = (I_0'/\beta) \cdot k_1' \int e^{-Kt} \cdot t ds, \tag{5}$$

$$I_B' = (I_0'/\beta) \cdot k_2' \int e^{-Kt}(t + \alpha t^n) ds, \tag{6}$$

where the integrations are carried out over the whole specimen. For aluminum it was found[2] that $n = 5$, and this value will be maintained in the following calculations.

Thin Films of Uniform Thickness

In this case

$$C' = \frac{k_1'}{k_2'} \cdot \frac{1}{(1 + \alpha t^4)}. \tag{7}$$

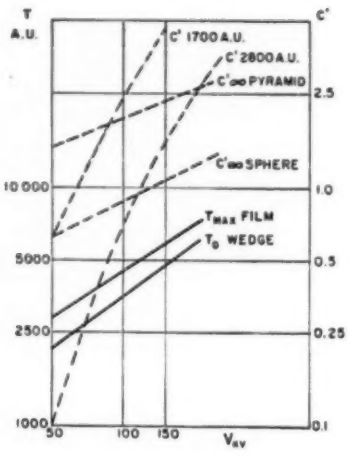


FIG. 2.

Pyramids

Suppose that the pyramids have a square base of side $2L$ perpendicular to the electron beam, and a height T. Then

$$ds=\frac{2L^2}{T}\left(1-\frac{t}{T}\right)dt, \tag{8}$$

$$I_R'=\frac{I_0'}{\beta}\cdot k_1'\cdot\frac{2L^2}{TK^2}\cdot\left[1-\frac{2}{TK}\right], \tag{9}$$

$$I_B'=\frac{I_0'}{\beta}\cdot k_2'\cdot\frac{2L^2}{T}$$

$$\times\left\{\frac{1}{K^2}\left(1-\frac{2}{TK}\right)+\alpha\left[\frac{120}{K^6}-e^{-KT}\cdot S_5(T)\right]\right.$$

$$\left.-\frac{\alpha}{T}\left[\frac{720}{K^7}-e^{-KT}\cdot S_6(T)\right]\right\}, \tag{10}$$

where

$$S_5(T)=\frac{T^5}{K}+\frac{5T^4}{K^2}+\frac{20T^3}{K^3}+\cdots, \tag{11}$$

$$S_6(T)=\frac{T^6}{K}+\frac{6T^5}{K^2}+\frac{30T^4}{K^3}+\cdots. \tag{12}$$

For T very large the contrast becomes

$$C'=\frac{k_1'}{k_2'}\cdot\frac{1}{[1+(120\alpha/K^4)]}. \tag{13}$$

Spheres

If the specimen consists of spherical particles of radius R,

$$ds=(\pi/4)\cdot t\,dt, \tag{14}$$

$$I_R'=\frac{I_0'}{\beta}\cdot k_1'\cdot\frac{\pi}{4}\cdot\frac{2}{K^3}, \tag{15}$$

$$I_B'=\frac{I_0'}{\beta}\cdot k_2'\cdot\frac{\pi}{4}\left\{\frac{2}{K^3}+\alpha\left[\frac{720}{K^7}-e^{-2KR}\cdot S_6(2R)\right]\right\}. \tag{16}$$

For R large,

$$C'=\frac{k_1'}{k_2'}\cdot\frac{1}{[1+(360\alpha/K^4)]}. \tag{17}$$

Truncated Wedges

If the wedge has a minimum thickness T_0, a very large maximum thickness, and

$$ds=\eta dt, \tag{18}$$

$$I_R'=\frac{\eta}{\beta}\cdot\int_{T_0}^{\infty}I_0'k_1'e^{-Kt}\cdot t\,dt \tag{19}$$

$$=\frac{\eta}{\beta}I_0'k_1'\frac{e^{-KT_0}}{K}\left(T_0+\frac{1}{K}\right), \tag{20}$$

$$I_B'=\frac{\eta}{\beta}\cdot I_0'k_2'\left\{\frac{e^{-KT_0}}{K}\left(T_0+\frac{1}{K}\right)+\alpha e^{-KT_0}S_5(T_0)\right\}, \tag{21}$$

and

$$C'=\frac{k_1'}{k_2'}\cdot\left(T_0+\frac{1}{K}\right)\Big/\left[\left(T_0+\frac{1}{K}\right)+\alpha K S_5(T_0)\right]. \tag{22}$$

By use of the experimental results[2] for aluminum, C' for 1700A and 2800A films has been plotted in Fig. 2 as a function of accelerating voltage. By use of Eqs. (13) and (17), C' for very large pyramids and spheres has also been plotted. These results apply to the 111 ring. It was found that when using Kodak Medium Lantern Slide Plates sufficient contrast to permit the measurement of ring diameters required that $C'\geqslant 0.1$. The experimental values for $T_{\max}$ film at which $C'=0.1$ are plotted in Fig. 2. Putting $C'=0.1$ in Eq. (22), the values of T_0 for a polycrystalline aluminum wedge have been calculated and are also shown.

DISCUSSION

The experimental results show that if the specimen thickness is nowhere very great compared with K^{-1}, then the contrast increases rapidly with increasing electron speed. The calculations for very large spheres and prisms indicate that if a major part of the specimen has a thickness large compared with K^{-1} the contrast increases slowly with increasing electron speed. The theory indicates that provided k_1'/k_2' is large enough the background due to the thick regions of the specimen will not obscure the rings, which are produced by the thinner regions. This is in agreement with general experience.

Electron diffraction by an infinite polycrystalline sphere is essentially reflection electron diffraction. The results of Fig. 2 suggest that if the specimen is smooth the contrast will increase approximately as $V^{\frac{1}{2}}$ for voltages in the range 50 to 150 kv. If the specimen is rough with thin projections (thickness of the order K^{-1}) the contrast will increase more rapidly with voltage. A further advantage may accrue from the use of higher accelerating voltages in reflection diffraction studies. This is a sharpening of the pattern and a consequent increase in peak contrast due to the reduction of the electron-optical refractive index $(1+V_i/V)^{\frac{1}{2}}$; V_i is the inner potential. While the effects of inner potential on peak contrast are automatically included in the measurement of k_1'/k_2' used here, these measurements are made on thin films in which the electrons probably enter the aluminum, on the average, closer to the normal than would be the case in reflection diffraction.

Experiments with a rough surface such as a heavy coating of magnesium oxide on copper and with a smooth surface such as an etched germanium single crystal suggest that the gain in contrast on raising the voltage is less marked in the case of the smooth surface.

The experimental values for $T_{\max}$, the maximum film thickness, and the calculated values of T_0, the minimum edge thickness of a wedge elsewhere very thick, which will yield a useful pattern suggest that these dimensions increase slowly with voltage. They increase approximately as $V^{\frac{1}{2}}$.

The presence of the term e^{-Kt} in Eq. (3) together with the fact that $K^{-1}=450$A for aluminum[2] at 50 kv indicates that for the specimen thicknesses commonly employed the ring intensity is not proportional to the mass of the diffracting crystals present in the part of the specimen irradiated. With two different substances present the ratio of ring intensities will only be proportional to the ratio of the masses and independent of the total mass of the specimen if the mean specimen thickness is small compared with K^{-1}. From this it may be concluded that the electron diffraction technique is not generally suitable for quantitative analysis.

CONCLUSIONS

An obvious conclusion from the previous section is that to obtain high contrast electron diffraction patterns the specimen thickness should be of the order K^{-1}. According to the theory outlined above, the contrast should approach k_1'/k_2' as the specimen thickness approaches zero. If the specimen is supported on a collodion film, this obviously will not be true. In any case if the reduction of specimen thickness results in a reduction of crystal size, the increased line broadening will reduce the peak contrast. In many practical cases where the contrast is poor a bigger increase in contrast can be obtained by a reduction in the mean specimen thickness than by an increase in accelerating voltage, at least in the range of voltages considered here.

It is nevertheless worthwhile to use the maximum available accelerating voltage since with all specimens an increase in contrast will result. The gain in range of specimen thicknesses that can be studied is more limited than might have been expected from the increase in contrast obtained when thin films are used. The final choice of voltage is therefore largely a matter of convenience. Using a single stage electron gun, 100 kv is approximately the highest accelerating voltage that can be employed without excessive breakdown, and at this voltage the patterns will still be large enough to measure accurately without the use of electronic magnification or the use of an inconveniently long camera. If higher voltages are to be considered, Möllenstedt's results[1] suggest that a major gain in the specimen thicknesses that can be examined will only result at voltages above 500 kv.

If the convenience of using an accelerating voltage around 50 kv is to be retained there are still three methods by which the contrast in electron diffraction patterns can be increased. These are

(a) by decreasing the background current I_B,
(b) by favoring the recording of the ring current I_R in comparison with the background current I_B,
(c) by using a higher contrast recording mechanism.

The first method has been used by Boersch.[3] The electrons that leave the specimen pass through a filter which turns back electrons that have lost more than a few electron-volts of energy. In consequence the background intensity is reduced without reduction in the ring intensity. Disadvantages of this scheme are that it requires the introduction of a high voltage lead into the camera and that the energy filter may cause distortions in the pattern, which would give trouble if highly precise measurements were to be made.

The second method has been used by v. Friesen[4] in some experiments on the energy losses of the background electrons. The scheme employed consists of a target film at which the electron diffraction pattern is formed and where the incident electrons generate x-rays. These x-rays pass through a filter to a photographic plate; the target, filter, and plate form a sandwich. The target material and accelerating voltage are chosen so that electrons that do not lose energy in the specimen will excite characteristic x-rays in the target, whereas those that lose a few electron volts of energy will not. The filter is intended to reduce the white radiation of frequency greater than the characteristic radiation. There is a considerable loss of speed of operation with this method due to recording x-rays rather than the electron charge directly. The divergence and scattering of the x-rays between the target and the photographic plate also leads to a loss of resolution.

Although it is possible to conceive of a number of

[3] H. Boersch, Optik **5**, 436 (1949).
[4] S. v. Freisen, Arkiv Mat. Astron. Fysik. **35A**, No. 18 (1948).

recording mechanisms based on the use of dc amplifiers and crystal counters, the photographic plate, because of its integrative character, is not likely to be supplanted from general use in the near future. The use of higher contrast emulsions would undoubtedly help in recording a given diffraction ring, and they are of value in recording single crystal patterns and Kickuchi lines. In a complete ring pattern there is usually a large range of current densities, and the use of a high contrast emulsion would require a series of exposures for different circular segments of the pattern. This scheme is not attractive for general use.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Methods of Measuring the Properties of Ionized Gases at High Frequencies. III. Measurement of Discharge Admittance and Electron Density*

DAVID J. ROSE† AND SANBORN C. BROWN
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 7, 1952)

Experimental methods are given for determining the complex admittance and electron density of gas discharges by the use of microwave techniques. Applications are discussed where the discharge is contained in either high or low Q resonant cavities.

I. INTRODUCTION

IN two previous papers,[1,2] referred to as Parts I and II, we have developed methods of measuring the Q and electric field in microwave cavities used for gas discharge work. This paper extends the techniques to include the properties of the discharge itself—in particular, of its complex admittance and electron density. As in the preceding papers, the most elegant methods apply to high Q cavities, and thus necessarily to low electron density discharges. However, at microwave frequencies, this restriction is not too severe for many purposes. Even if the discharge is not maintained in a high Q cavity and the electron density is high, some analysis can be made on the problem; this matter is also taken up.

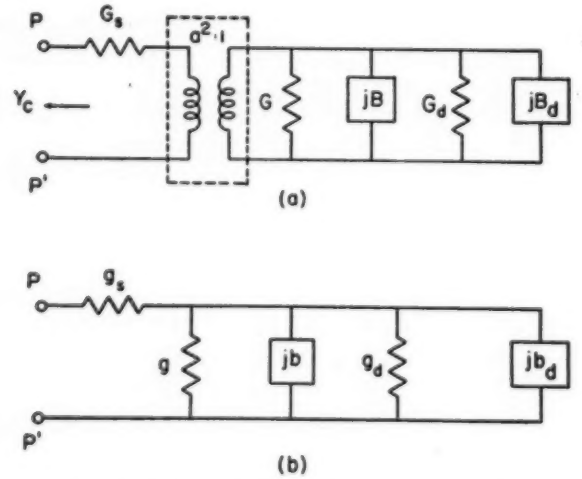


FIG. 1. Equivalent circuits for a resonant cavity containing a discharge.

II. DISCHARGE ADMITTANCE BY CAVITY METHODS

If the cavity Q is high and the presence of the discharge does not affect its electrical properties unduly, the discharge admittance can be measured easily.

The cavity is represented by the equivalent circuit of Fig. 1(a), and the transmission line is coupled to it at PP'. The discharge is represented by the conductance G_d and susceptance jB_d. This circuit can be replaced by that of Fig. 1(b), wherein the transformer is removed and the line admittance is normalized to unity. A detailed discussion of this general equivalence is given in Part I. Here

$$\frac{g_s}{G_s}=\frac{1}{Y_c}, \quad \frac{g}{G}=\frac{b}{B}=\frac{g_d}{G_d}=\frac{b_d}{B_d}=\frac{a^2}{Y_c}. \tag{1}$$

The susceptance jb is given by

$$b=\beta(\omega/\omega_0-\omega_0/\omega), \tag{2}$$

where β is a coupling coefficient between the cavity and transmission line and is described in Part I. The angular resonant frequency of the empty cavity is ω_0.

Inasmuch as the empty cavity resonance is given by $b=0$, or $\omega=\omega_0$, it is apparent that the presence of the discharge susceptance jb_d alters the resonant frequency to a new value ω_0' given by

$$\beta(\omega_0'/\omega_0-\omega_0/\omega_0')+b_d=0. \tag{3}$$

The entire equivalent circuit analogy is valid only as long as the presence of the discharge perturbs the frequency slightly; that is, $\omega_0'-\omega_0 \ll \omega_0$. Thus

$$b_d \approx -2\beta(\omega_0'-\omega_0)/\omega_0; \tag{4}$$

this shift in the cavity resonance when the discharge is initiated can be measured. The value of β is determined as described in Part I. In this manner the discharge susceptance is determined.

At the same time, the discharge conductivity has the same effect as changing the cavity losses to g', where

$$g'=g+g_d. \tag{5}$$

* This work has been supported in part by the Signal Corps, Air Materiel Command, and ONR.

† Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] S. C. Brown and D. J. Rose, J. Appl. Phys. 23, 711 (1952).

[2] D. J. Rose and S. C. Brown, J. Appl. Phys. 23, 719 (1952).

It is shown in Part I that the quantity $g_0=g'/(1+g'/g_s)$ determines the standing-wave ratio at resonance. Actually this was shown for the special case $g_d=0$, but since the conductances add, the present result follows simply. Since g_s determines the off-resonant standing-wave ratio, g' can be determined from these two quantities. In Fig. 2, g' is plotted as a function of the standing-wave ratio at resonance R_0 (decibels) for several values of R_∞. Negative values of R_0 indicate undercoupling, and positive values overcoupling. R_0 is to be measured at ω_0 for the empty cavity and at ω_0' when it contains the discharge. R_∞ is the same for both conditions. By measuring R_0 at ω_0 for the empty cavity, $g'\equiv g$ is determined from Fig. 2. With the discharge present, R_0' at ω_0' is measured. Again $g'=g+g_d$ is determined from Fig. 2, and g_d is obtained by subtraction. The discharge will always tend to undercouple the cavity; for this reason it is often advisable to overcouple it, when empty, by 10 or 12 db so that when the discharge is introduced, the cavity will not be too greatly undercoupled. In order to keep the discharge constant in the cavity as the necessary frequency excursions are made, the light output of the discharge can be monitored. This method is described in Part I.

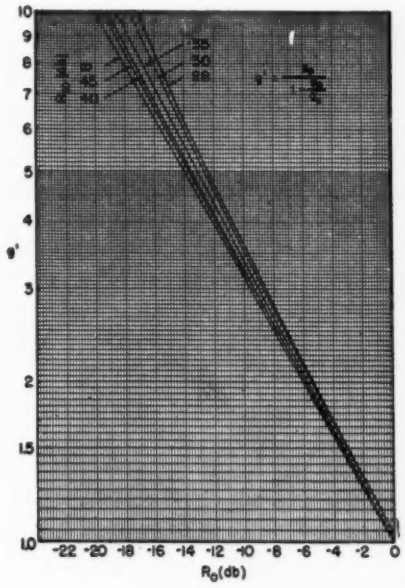
In addition to the restriction on frequency perturbation, it is essential that the cavity Q remain high with a discharge present. These two restrictions mean that the absolute magnitude of the electron current must be much smaller than the displacement current, or that the conductivity of the discharge $|\sigma|\ll\omega\epsilon_0$. Even so, at microwave frequencies, electron densities of the order of $10^{10}/\text{cm}^3$ in a gas at 20 mm Hg pressure can be handled by these techniques.

Occasionally, where the electric field configuration in the cavity is simple, it is convenient to speak of voltages in the cavity and to calculate the quantities G_d and jB_d not normalized. These are obtained from Eq. (1) after calculating the transformer ratio a^2. First, it is necessary to define a voltage across the cavity as $V=\int\text{Eds}$ along the desired path. In Part II we defined a quantity η which related the stored energy and electric field, and this in a sense by-passed the calculation of the transformer ratio. It is possible, however, to calculate a capacitance C from η relating the energy and voltage. The normalized capacitance c associated with the circuit is $c=\beta/\omega_0$; thus the ratio $a^2=cY_c/C$. From the definition of η and consideration of the stored energy

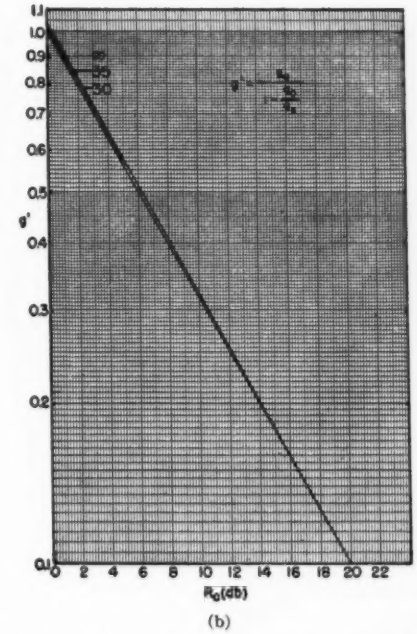
$$C=\eta E_\eta^2\Big/\Big[\int\text{Eds}\Big]^2,$$

where E_η is the field at the reference point chosen for the calculation of E through the use of η. Thus

$$a^2=\beta Y_c\Big[\int\text{Eds}/E_\eta\Big]^2\Big/\omega_0. \tag{7}$$



(a)



(b)

FIG. 2. The conductance g' plotted as a function of R_0 and R_∞.

III. CONDUCTIVITY AND ELECTRON DENSITY BY CAVITY METHODS

If cavity methods apply, the discharge admittance is given by[3]

$$g_d+jb_d=\frac{\beta\int\sigma E^2dv}{\omega_0\epsilon_0\int E^2dv}, \tag{8}$$

[3] J. C. Slater, Revs. Modern Phys. 18, 481 (1946).

where $\sigma=\sigma_r+j\sigma_i=\mathbf{J}/\mathbf{E}$ is the complex conductivity of the ionized plasma, and the integrals are to be taken over the whole volume containing the field; $\mathbf{J}$ is the current density. Since the electrons are very much more mobile than the ions, the conductivity at high frequencies is due solely to their presence. The imaginary part $j\sigma_i$ of σ gives rise to the discharge susceptance jb_d and the shift in cavity resonance. The real part σ_r lowers the Q through the conductance g_d. Either of these phenomena can be used for the measurement of the electron density.

Electron Density by Measurement of b_d

The complex conductivity of the discharge is dependent on the angular frequency ω and on the collision frequency ν_c of the electrons with gas atoms. ν_c is proportional to the gas pressure; at low pressure, $\nu_c \ll \omega$ and σ_i is independent of ν_c and is

$$\sigma_i=-ne^2/m\omega, \tag{9}$$

where e and m are the electronic charge and mass, and n is the electron density. Thus

$$b_d=\frac{-\beta e^2\int nE^2dv}{m\epsilon_0\omega_0^2\int E^2dv}. \tag{10}$$

b_d and the other constants can be measured. The integrals can be performed for any simple geometrical configuration of E and n. For example, for a circular cylindrical cavity in the TM_{010}-mode, $E_r=E_\phi=0$, and $E_z\sim J_0(2.405r/r_0)$, where r and z are radial and axial dimensions, and r_0 is the cavity radius. If the discharge is controlled by diffusion, $n=n_0\sin(\pi z/h)J_0(2.405r/r_0)$; n_0 is the density at the center. In this case, Eqs. (4) and (10) yield

$$n_0=4.87\times10^{13}\Delta\lambda/\lambda^3\ \text{cm}^{-3} \tag{12}$$

in cgs units, where $\Delta\lambda$ is the shift in the resonant wavelength. Solutions for various distributions of n within a TM_{010}-mode cavity are given in Table I.

If ν_c is not much less than ω, the shift in resonance is a function of ν_c/ω also, and the calculations generally become involved. If ν_c is independent of electron energy, σ is given by[4]

$$\sigma=\frac{ne^2(\nu_c-j\omega)}{m(\nu_c^2+\omega^2)}. \tag{13}$$

The imaginary part should be compared with Eq. (9). The integrals over the cavity are the same, and for the mode discussed above

$$n_0=4.87\times10^{13}[1+(\nu_c/\omega)^2]\Delta\lambda/\lambda^3\ \text{cm}^{-3}. \tag{12a}$$

If ν_c is not independent of electron energy, some knowledge both of its dependence and of the electron velocity distribution is required; then σ is calculated by kinetic formulas.

There are various applications of this method. The discharge might be maintained by some means in a glass container. Then a microwave cavity could be fitted around the container, and the microwave field used as a low level probe to determine the density. Or again, the microwave power might be used to maintain the discharge entirely, and the wavelength of this power could be varied to search out the new resonance of the cavity with the discharge. Measurements of electron density in the thermal afterglow of a pulsed discharge have been made by Biondi and Brown[5] and others. They used a low level probe to determine the frequency shift of the cavity as a function of time after the start of the afterglow. These latter measurements were made by determining the time of appearance of a minimum reflected signal on the input line as a function of frequency shift. They can also be made by determining the time of maximum transmission through the cavity as a function of frequency shift, by attaching a small output coupling to the cavity.

TABLE I. Electron density in a container placed in the center of a cylindrical TM_{010} cavity as a function of change of resonant wavelength. Cavity radius $=r'$, and height $=H$.

Container shape	Electron distribution	Formula	Remarks
Cylindrical radius r_0 height h	uniform $=n$	$n=\frac{H}{h}\frac{\Delta\lambda}{\lambda^3}\frac{6.04\times10^{12}}{\rho^2[J_0^2(2.4\rho)+J_1^2(2.4\rho)]}\ \text{cm}^{-3}$	$\rho=r_0/r'$ electron distribution characteristic of recombination loss
	$n=n_0J_0\left(\frac{2.4r}{r_0}\right)\cos\left(\frac{\pi z}{h}\right)$	$n_0=\frac{H}{h}\frac{\Delta\lambda}{\lambda^3}\frac{3.51\times10^{13}}{[1.6\rho^2-1.42\rho^4+2.58\rho^6]}\ \text{cm}^{-3}$	accurate for $\rho\leqslant\frac{1}{2}$ electron distribution characteristic of diffusion loss
any	any with average electron density $=\bar{n}$	$\bar{n}=\frac{6.04\times10^{12}\Delta\lambda\ (\text{volume of cavity})}{\lambda^3\ (\text{volume of container})}\ \text{cm}^{-3}$	accurate for largest dimension in radial direction $\ll r'$

[4] E. Everhart, Phys. Rev. **76**, 839 (1949).
[5] M. A. Biondi and S. C. Brown, Phys. Rev. **75**, 1700 (1949).

Electron Density by Measurement of g_d

If the discharge conductivity is to be used to measure electron density, the dependence of ν_c on electron energy must be taken into account. If ν_c is independent of energy, Eqs. (8) and (13) yield

$$g_d=\{\beta e^2(\nu_c/\omega)/\epsilon_0\omega_0^2 m[1+(\nu_c/\omega)^2]\} \times\left\{\int nE^2dv \Big/ \int E^2dv\right\}. \quad (14)$$

For the TM_{010}-mode treated above,

$$n_0=2.44\times10^{13}[1+(\nu_c/\omega)^2]g_d/\beta\lambda^2(\nu_c/\omega)\ \text{cm}^{-3} \quad (15)$$

in cgs units; g_d is determined from Fig. 2 as described above.

Determination of Probability of Collision

From Eq. (8), the ratio $\sigma_r/\sigma_i=g_d/b_d$, and this ratio can be measured. If ν_c is independent of electron energy, then $\nu_c/\omega=g_d/b_d$ and ν_c can be determined simply. If ν_c is not constant, but some assumption about its velocity dependence is made, the ratio σ_r/σ_i can be calculated. Since $\nu_c=vpP_c$, where v is the electron velocity, p the gas pressure, and P_c the probability of collision, P_c can be measured in this manner. Phelps, Fundingsland, and Brown[6] have used this technique to measure P_c for thermal electrons in the afterglow of a pulsed discharge. Using a low level probe and pulse techniques, they measured g_d and b_d simultaneously at specific times in the afterglow.

IV. COMPLEX ADMITTANCE AND CONDUCTIVITY AT HIGH ELECTRON DENSITIES

If the cavity Q is low or the electron density is high so that $|\sigma|$ is comparable with $\omega\epsilon_0$, it is necessary to use the four terminal network method described in Part II, and impedance measurements on the transmission line will determine the discharge parameters G_d and B_d. If such methods are to apply, it is necessary that the discharge region be small compared to the wavelength. In this section, we shall investigate this region to show how G_d and B_d might be calculated from the conductivity.

It will be necessary to consider both the real and displacement currents. According to our convention of characterizing the ionized medium by a complex conductivity σ, Maxwell's equation $\nabla\times\mathbf{H}=\dot{\mathbf{D}}+\mathbf{J}$ becomes

$$\nabla\times\mathbf{H}=(j\omega\epsilon_0+\sigma_r+j\sigma_i)\mathbf{E}; \quad (16)$$

the term in brackets may be looked upon as a total admittance of the region. An alternative and equivalent viewpoint is to characterize the medium by a complex permittivity $\epsilon=\epsilon_r-j\epsilon_i$ in which the effect of the real conductivity is included in ϵ_i. According to this viewpoint, Eq. (16) becomes

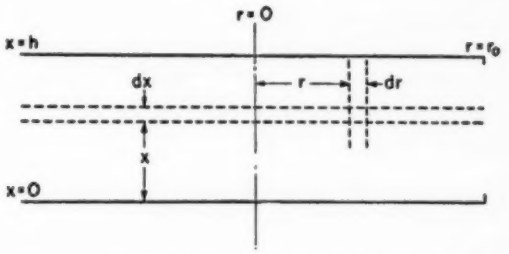


FIG. 3. Cross section of a gas discharge between parallel plates.

$$\nabla\times\mathbf{H}=j\omega(\epsilon_r-j\epsilon_i)\mathbf{E}=[j\omega\epsilon_0+\omega\epsilon_i+j\omega(\epsilon_r-\epsilon_0)]\mathbf{E}. \quad (17)$$

Comparison of these two equations shows that

$$\sigma=\omega\epsilon_i+j\omega(\epsilon_r-\epsilon_0). \quad (18)$$

One system may be converted to the other in this manner.

The calculation of G_d and B_d from σ_r and σ_i is difficult, and depends on the spatial distribution of the discharge; therefore, in practice, the simplest possible discharge configuration is chosen. An example will be given.

Suppose the discharge exists between circular parallel plates of radius r_0 and spacing $h\ll\lambda$, as in Fig. 3. The electron concentration will in general be a function of both x and r; but for simplicity assume the concentration to be independent of r for $0\leq r\leq r_0$, and to be zero for $r>r_0$. Since $h\ll\lambda$, the current will be continuous across the gap if both $\mathbf{J}$ and $\mathbf{D}$ are included. Fringing fields will be neglected, which means that the equipotentials are parallel. Under these circumstances, the following computation is justified.

The admittance associated with a slab of thickness dx is, from Eq. (16), $\pi r_0^2\{\sigma r(x)+j[\sigma_i(x)+\pi\epsilon_0]\}/dx$. The impedance of the slab is the reciprocal of this admittance, so that the impedance of the whole gap is

$$Z=\int_0^h \frac{dx}{\pi r_0^2\{\sigma_r(x)+j[\sigma_i(x)+\omega\epsilon_0]\}}. \quad (19)$$

This impedance is not quite the reciprocal of the admittance of the discharge as determined by impedance measurements. Experimentally, a susceptance B_c arising from the capacity of the empty plates was automatically subtracted out where

$$B_c=\pi r_0^2\omega\epsilon 0/h. \quad (20)$$

Thus

$$\frac{1}{G_d+j(B_d+B_c)}=\int_0^h \frac{dx}{\pi r_0^2\{\sigma_r(x)+j[\sigma_i(x)+\omega\epsilon_0]\}} \quad (21)$$

gives the relation between the measured values of G_d, B_d, and σ_r, σ_i. The appropriate form of σ must be determined from kinetic theory.

If Eq. (21) is separated into real and imaginary parts, in the limit of low electron densities where $|\sigma|\ll\omega\epsilon_0$,

[6] For example, see Phelps, Fundingsland, and Brown, Phys. Rev. 84, 559 (1951).

$|G_d+jB_d| \ll B_e$, and

$$G_d=\frac{\pi r_0^2}{h}\int_0^h \sigma_r(x)dx \tag{22}$$

$$B_d=\frac{\pi r_0^2}{h}\int_0^h \sigma_i(x)dx. \tag{23}$$

Thus at low densities G_d is a function of σ_r, and B_d is a function of σ_i, as was shown more generally in the earlier sections of this paper. If $|\sigma|$ is comparable to or larger than $\omega\epsilon_0$, however, no such simple relations apply, and calculations similar to those leading to Eq. (21) must be made to suit the actual experimental arrangement.

The calculation of electron density by this method is more difficult than by cavity methods, and we can proceed by assuming a spatial distribution and a form for the conductivity σ. The necessary integrations over the discharge region are then performed. The calculated values of G_d and B_d are compared with the experimental ones, and agreement indicates the correctness of the assumptions.

The authors wish to acknowledge the assistance of Dr. Melvin A. Herlin, Dr. Edgar Everhart, and Dr. Manfred A. Biondi for their contributions to Massachusetts Institute of Technology Research Laboratory of Electronics Technical Reports 66 and 140. Some of the techniques of this paper were based on those presented in these earlier publications.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Self-Diffusion in Pure Polycrystalline Silver*

L. Slifkin, D. Lazarus, and T. Tomizuka†
Department of Physics, University of Illinois, Urbana, Illinois
(Received May 14, 1952)

Self-diffusion in pure polycrystalline silver has been measured using Ag^{110} as a tracer. The results of the high temperature experiments, where volume diffusion is predominant, are in excellent agreement with those of other investigators. A least-squares calculation using all available data on volume diffusion of silver gives as a value for the diffusion coefficient $D=0.724e^{-45,500/RT}$. The occurrence at low temperatures of grain boundary diffusion, in which the activity decreases exponentially with the first power of the penetration depth, as reported by Hoffman and Turnbull, is confirmed.

INTRODUCTION

An adequate understanding of the mechanisms involved in diffusion in solids rests in a large measure on the accuracy of the available data. A number of difficulties involved in generalizing the results of some early experiments was recently emphasized by Nowick.[1] In particular, it is important experimentally to separate those results due primarily to diffusion through the bulk of a material from those associated with diffusion along grain boundaries or other short-circuiting paths. The recent work of Hoffman and Turnbull[2] and Fisher[3] indicate that different solutions for the diffusion equation are appropriate for the two cases.

The present investigation was undertaken as a preliminary experiment in a set of diffusion measurements with silver solid solutions in an attempt to verify and extend these previous results.

EXPERIMENTAL

Handy and Harmon 99.99 percent silver was cast into cylinders of $\frac{3}{4}$-in. diameter under an atmosphere of helium. The crucibles used were of high purity graphite (National Carbon Company grade AUC) which were cleaned by boiling in aqua regia, leaching in distilled water, and baking in vacuum at 900°C. The silver cylinders were alternately deformed in compression and annealed at 625°C in vacuum until a uniform grain size of $\frac{1}{2}$ to 1 mm was obtained. They were then cut and machined to form drum-shaped specimens of $\frac{5}{8}$-in. diameter and about $\frac{3}{8}$-in. length, after which they were annealed at 625°C. One face of each specimen was polished and etched with dilute HNO_3.

A layer of radioactive silver, of about 10A thickness, was plated onto the polished face of the specimen. This layer contained about 0.5 μc of Ag^{110}. During plating, the sides of the specimen were protected by a coating of paraffin or beeswax. The plating solution was a cyanide bath in which all constituents were at quite low concentrations, to avoid adding any inactive silver to the high specific activity material obtained from Oak Ridge. The current density was 8 ma/dm^2.

The plated specimens were sealed off in evacuated Vycor tubes and placed in furnaces to diffuse. Within the furnace, each specimen resided inside a massive nichrome block. This block insured a uniform temperature within the cavity and, by its thermal inertia, smoothed out temperature fluctuations. The furnace operating at temperatures above 750°C was controlled

* This work was supported in part by the AEC.
† On leave from the Institute of Science and Technology, Tokyo University.
[1] A. S. Nowick, J. Appl. Phys. **22**, 1182 (1951).
[2] R. E. Hoffman and D. Turnbull, J. Appl. Phys. **22**, 634 (1951).
[3] J. C. Fisher, J. Appl. Phys. **22**, 74 (1951).

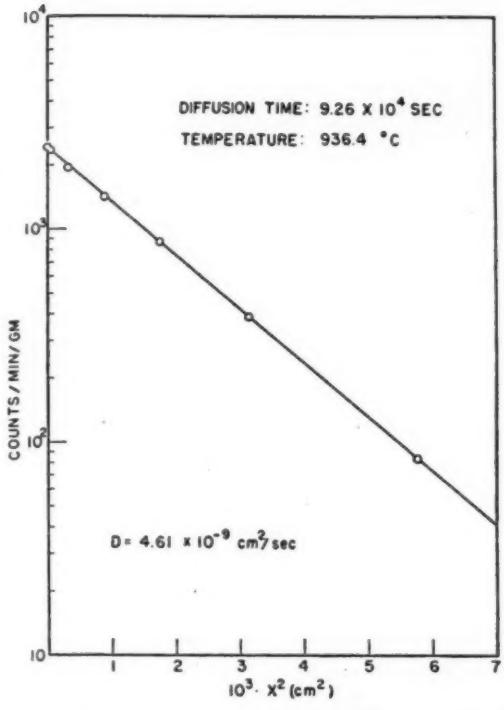


FIG. 1. Penetration curve for volume diffusion at 936°C.

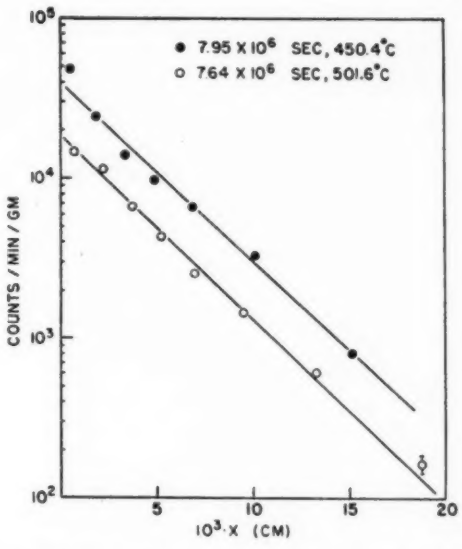


FIG. 2. Penetration curves for grain boundary diffusion. The upper curve has been displaced upward by a factor of 3. The probable error shown for one of the points was a counting error and applies to that point only.

by a Tag Celectray, and the temperature at the specimen was maintained constant within limits of ±1°C. Those furnaces operating at lower temperatures were controlled by a simple bridge type controller[4] employing a platinum resistance winding near the heater as one arm of a Wheatstone bridge. In these furnaces the temperature at the specimen was constant to within ±0.6°C. Temperatures were measured with chromel-alumel thermocouples which had been checked against a calibrated *Pt—Pt, Rh* couple.

TABLE I. Coefficients of volume self-diffusion in silver.

Temp. (°C)	D (cm²/sec)	Specimen	Investigators
872	1.37×10^{-9}	Single crystal	Hoffman and Turnbull
876	1.74×10^{-9}	Single crystal	Hoffman and Turnbull
870	1.33×10^{-9}	Single crystal	Hoffman and Turnbull
819	5.85×10^{-10}	Single crystal	Hoffman and Turnbull
798	3.54×10^{-10}	Polycrystal	Hoffman and Turnbull
798	4.01×10^{-10}	Single crystal	Hoffman and Turnbull
604	2.84×10^{-12}	Single crystal	Hoffman and Turnbull
604	3.04×10^{-12}	Single crystal	Hoffman and Turnbull
501	2.92×10^{-13}	Single crystal	Hoffman and Turnbull
501	1.58×10^{-13}	Single crystal	Hoffman and Turnbull
500	1.68×10^{-13}	Single crystal	Hoffman and Turnbull
500	2.28×10^{-13}	Single crystal	Hoffman and Turnbull
948	5.20×10^{-9}	Polycrystal	W. A. Johnson
876	1.55×10^{-9}	Polycrystal	W. A. Johnson
802	3.97×10^{-10}	Polycrystal	W. A. Johnson
726	7.80×10^{-11}	Polycrystal	W. A. Johnson
936	4.61×10^{-9}	Polycrystal	Present authors
794	3.64×10^{-10}	Polycrystal	Present authors
666	2.45×10^{-11}	Polycrystal	Present authors

[4] J. M. Wilson, Electronics **23**, 84 (1951).

After the diffusion anneal, the specimen was mounted on a precision lathe and 0.003–0.020 in. was turned off from the side to eliminate effects of surface diffusion. (The depth removed in each case depended upon the extent to which diffusion was expected to have taken place.) The specimen was then sectioned. The lathe collet was mounted in a special chuck which permitted the active face of the specimen to be oriented accurately perpendicular to the lathe axis. The material from each cut was collected in a specially designed shield and weighed in an annular glass vessel. Dilute HNO_3 was added to dissolve the silver. The vessel was then placed over a gamma-sensitive Geiger counter to determine the Ag^{110} concentration in the cut.

RESULTS

The self-diffusion of silver was measured at six temperatures over the range 450°–936°C. The results of the diffusion anneals at the three higher temperatures show that the activity falls off exponentially as the square of the penetration depth. A typical result is shown in Fig. 1. The diffusion coefficients obtained from these data are given in Table I.

The data from the anneals at 501.6°C and 450.4°C are shown in Fig. 2. At these temperatures the effects of grain boundary diffusion predominate, and the activity decreases exponentially with the first power of the penetration depth. The transition between volume and grain boundary diffusion at 576°C is shown in Fig. 3. For small penetration depth the logarithm of the activity depends on the square of the depth, but for larger depth this quantity varies approximately linearly with depth.

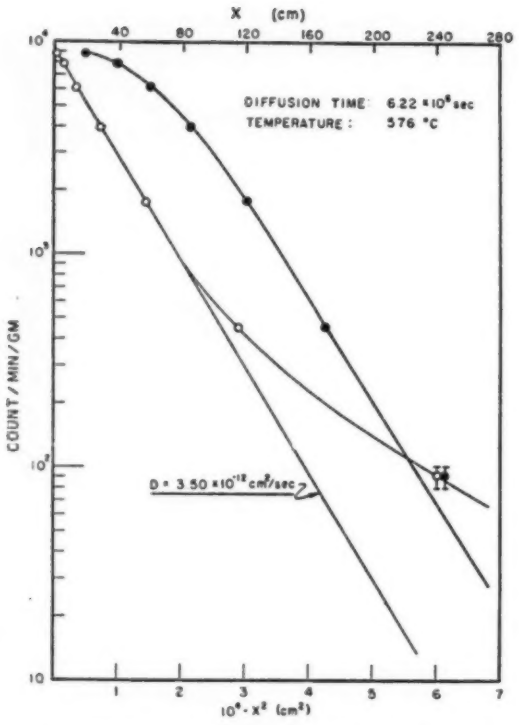


FIG. 3. Penetration curve showing combined effects of volume and grain boundary diffusion. The top scale for the abscissa applies to the closed circles, the bottom scale to the open circles. The probable error shown for the last point was a counting error and applies to that point only.

DISCUSSION

In the diffusion curves for polycrystals of Hoffman and Turnbull[2] an unusually high activity was noted at small penetration depths. Similar results have been obtained by Fensham[5] in studing diffusion in tin single crystals and by Miller and Banks[6] with zinc. Fensham has proposed that this anomaly is due to a non-equilibrium distribution of vacancies resulting in an increase in the diffusion rate near the surface.

In the present investigation the first point of the penetration curve was never significantly removed from the extrapolation of the succeeding points either for volume diffusion or for grain boundary diffusion (see Figs. 1, 2, and 3). The radio-silver available was of such high specific activity that very thin electroplates (about 10A) could be employed. The anomaly in the previous work may therefore be due simply to the greater thickness of the electroplated layers used.

All the available data on volume self-diffusion of silver are given in Table I. Included are the results of W. A. Johnson,[7] those of Hoffman and Turnbull[8] on single- and polycrystals, and the present work. These data are plotted in Fig. 4 as log diffusion coefficient *versus* reciprocal absolute temperature. The close agreement between the work of the three different laboratories is very encouraging and demonstrates the feasibility of obtaining reproducible diffusion data. The line drawn through the points was determined by a least-squares treatment of all the data except the 500°C results of Hoffman and Turnbull. These low temperature data on single crystals were omitted because they were thought to be considerably less reliable than the high temperature measurements.[8] The intercept and slope of the line give a value for the diffusion coefficient $D=0.724 \exp(-45{,}500/RT)$ cm²/sec.

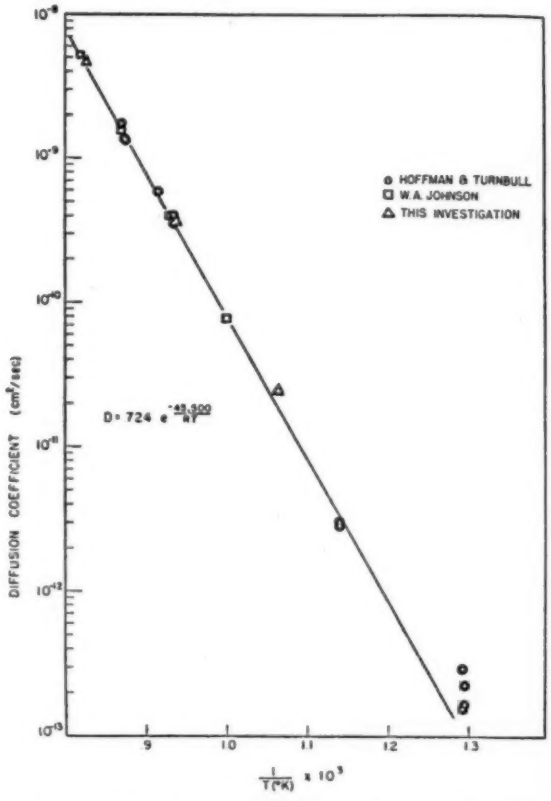


FIG. 4. Composite plot of volume diffusion coefficients of silver in silver.

The results on grain boundary diffusion closely check those of Hoffman and Turnbull. It should be noted that whereas these workers found grain boundary effects at temperatures as high as 660°–700°C, the present results show the onset of grain boundary diffusion to occur at temperatures about 100° lower. The difference most probably results from the larger grain size used in the present measurements.

Following the analysis of Fisher,[3] the following values for the grain boundary diffusion coefficient were obtained from the data of Fig. 2: 3.8×10^{-8} cm²/sec (501.6°C) and 1.3×10^{-8} cm²/sec (450.4°C). These data are consistent with Hoffman and Turnbull's value of 20 kcal/mole for the activation energy for grain boundary diffusion.

[5] P. J. Fensham, Australian J. Sci. Research **A3**, 91 (1950).
[6] P. H. Miller and F. R. Banks, Phys. Rev. **61**, 648 (1942).
[7] W. A. Johnson, Trans. Am. Inst. Mining Met. Engrs. **143**, 107 (1941). Johnson's diffusion coefficients, as plotted in his paper, were modified to take account of lattice expansion from room temperature (at which the measurements of depth were made) to the temperature at which diffusion took place. For comparison with the other results, his data were normalized to room temperature. The correction is small (of the order of 1 percent).
[8] R. E. Hoffman (private communication).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

Reliability of Probe Measurements in Hot Cathode Gas Diodes

GOTTFRIED WEHNER AND GUSTAV MEDICUS
Components and Systems Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio
(Received June 2, 1952)

Experimental investigations show that Langmuir probe measurements can give quite erroneous results, due to work function changes of the probe, insulating layers on the probe surface, or different work functions at different places of the probe. The deposition of evaporated cathode compounds, the adsorption of oxygen released from oxide cathodes, and the sputtering and evaporation from the probe may change the work function of a pure tungsten probe in time scales of seconds. In oxide cathode tubes especially, it is necessary to measure immediately after the probe has been thoroughly cleaned. An $X-Y$ plotter, which recorded the whole probe characteristic in seconds, was found to be most suitable. No deviation from the Maxwellian velocity distribution of the plasma electrons could be found over a range of nearly four orders of magnitude in probe current, even in oxide cathode tubes under conditions where such deviations have been found by other authors.

INTRODUCTION

IN studies of the discharge mechanism in low voltage gas discharges, which may be maintained with voltage drops much below the lowest excitation potential of the gas,[1–4] probe measurements throughout the discharge are most important. Measuring plasma potentials of a few volts requires more accuracy in the determination of the actual surface potential of the probe than is normally used. This potential differs from the measured voltage on account of the work function of the probe or high resistance layers on the probe surface. The surface conditions of the probe can change so fast that the work function is not constant while the probe characteristic is recorded, especially when a static, point by point measuring method is used. This may escape notice, because the characteristic in many cases still can be reproduced when recorded upwards and downwards if at each point sufficient time is provided for reaching equilibrium between the different influences which affect the work function, such as adsorption of impurities, deposition of cathode material on the probe, and evaporation and sputtering from the probe. A changing work function of the probe not only affects the reliability of plasma potential measurements in low voltage discharges, but is of more general interest because the plasma electrons may be collected at different parts of the probe characteristic with a different probe work function, which may result in a faulty measurement of the plasma electron velocity distribution.

Van Berkel[5] seems to have been the first one to call attention to the fact that changing work functions may influence the reliability of probe measurements. Anderson[6] studied these changes experimentally by taking oscillographic pictures of the probe characteristic with varied frequency. In a rare gas discharge with tungsten cathode no changes appeared with varied frequency. However, in an oxygen discharge the recording frequency had to be raised to several hundred cycles in order to eliminate changes of the probe surface during one cycle. In a very recent publication, Easley[7] made probe measurements in oxide cathode Hg-A and Hg-Kr discharges. When she eliminated the influence of the changing work function of the probe by measuring each point of the probe characteristic immediately after the probe was cleaned, and when no oscillations, striations, or impurities were present, no deviations from the Maxwellian velocity distribution were found. However, when the probe was not properly cleaned, deviations from the straight line in the semilog plot appeared in a direction which could be explained by a higher work function of the probe at low probe temperature. She ascribes these results to an increase of the work function of a pure tungsten surface by contamination with mercury. In tubes with Ba cathodes or Ba-Sr-oxide cathodes, on the other hand, deviations from the Maxwellian velocity distribution are reported[2,8] which, if they are not real, could analogously be explained by a lower work function of the probe at low temperature. Since most of the earlier publications which report deviations from the Maxwellian velocity distribution do not mention the cleaning procedure of the probe, serious doubts concerning the reliability of these measurements may arise.

The principal objectives of our study were as follows:

1. To get reliable probe characteristics in the low voltage, hot cathode arc.
2. To check if the reported deviations from the Maxwellian velocity distribution in oxide cathode tubes can be confirmed.
3. To investigate the different influences which determine the work function of a probe in connection with the low voltage arc.

The principles and the theory involved in Langmuir probe measurements are well enough known so that they need not be discussed here. Good summaries and bibliographies of the many publications which followed Langmuir's basic work up to the year 1939

[1] K. T. Compton and C. Eckart, Phys. Rev. **25**, 129 (1925).
[2] M. J. Druyvesteyn, Z. Physik. **64**, 781 (1930).
[3] M. J. Druyvesteyn and F. M. Penning, Revs. Modern Phys. **12**, 87 (1940).
[4] Malter, Johnson, and Webster, RCA Rev. **12**, 415 (1951).
[5] Van Berkel, Physica **V**, 230 (1938).
[6] Anderson, Phil. Mag. **38**, 179 (1947).
[7] M. A. Easley, J. Appl. Phys. **22**, 590 (1951).
[8] R. Seeliger and R. Hirchert, Ann. Physik **11**, 817 (1931).

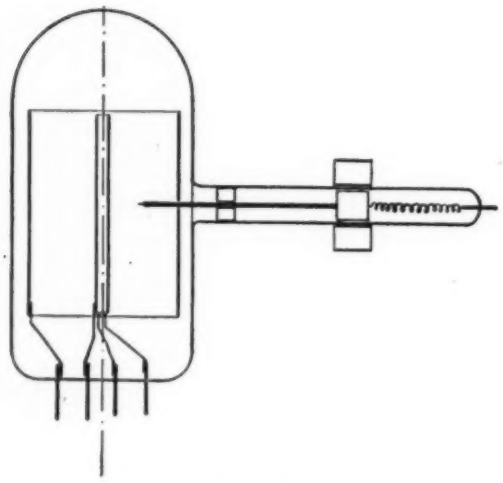


FIG. 1. Discharge tube No. 1.

are to be found in Loeb's book[9] and in an article by Rompe and Steenbeck.[10]

EXPERIMENTAL SET-UP

The Discharge Tubes

The measurements were made in three different tubes. Tube No. 1 was a demountable diode with concentric electrode arrangement (anode diameter 5 cm), and a probe protruding through a hole in the anode which could be moved magnetically between anode and cathode, as is shown in Figs. 1 and 2. The upper part of the tube could be easily separated from the cathode part by melting a lead joint under hydrogen atmosphere with an induction heater. Tube No. 2 was a sealed-off glass tube with a small cathode and a big plane anode (distance 3.6 cm), as shown in Fig. 3. The probe hung like a pendulum on a flexible support of

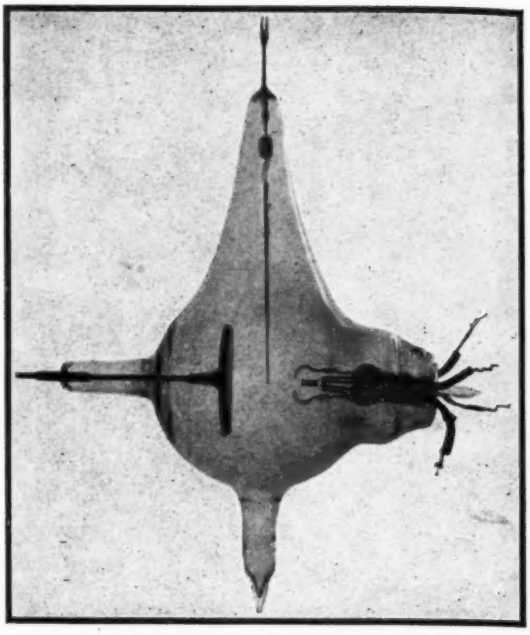


FIG. 3. View of discharge tube No. 2. Near cathode are two Ba-core nickel wires for the evaporation of barium onto the cathode.

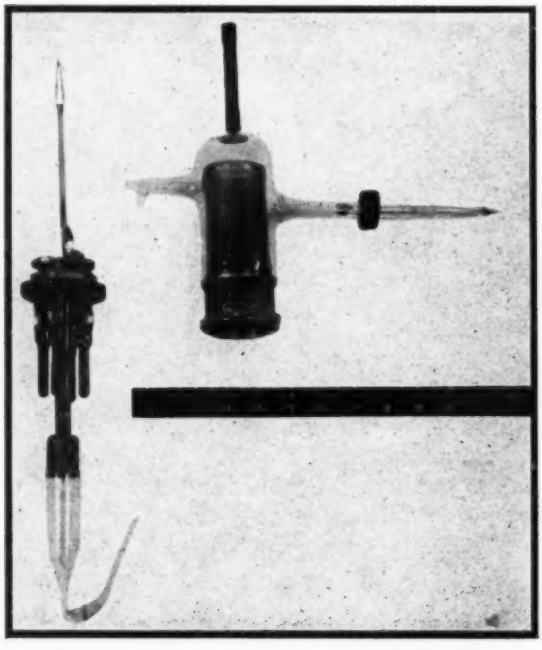


FIG. 2. View of disassembled tube No. 1. Scale in inches.

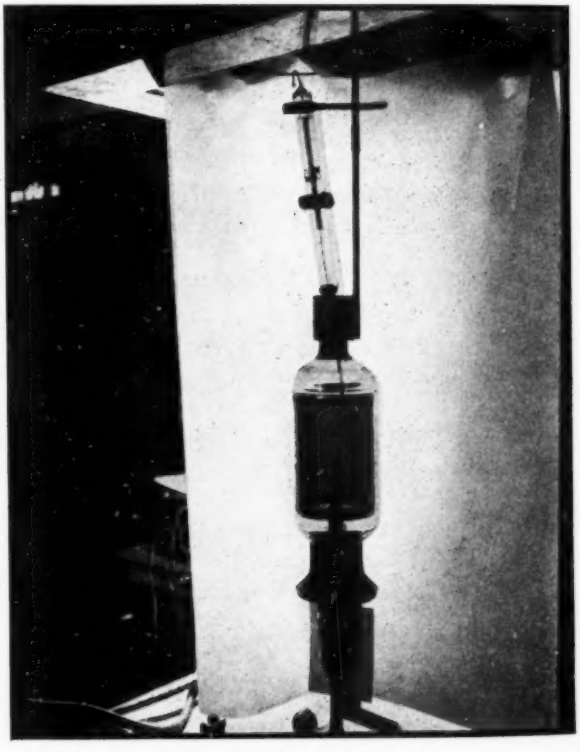


FIG. 4. View of discharge tube No. 3.

[9] Leonard B. Loeb, *Fundamental Processes of Electrical Discharges in Gases* (New York, 1939), p. 230.

[10] R. Rompe and M. Steenbeck, Ergeb. exakt. Naturwiss. **18**, 257 (1939).

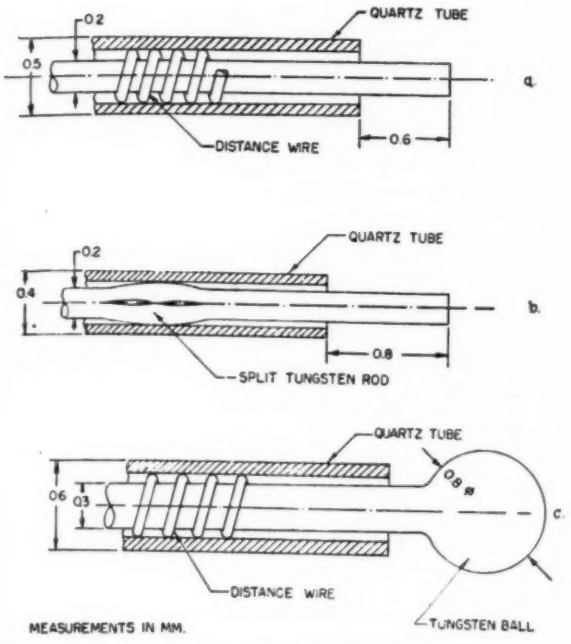


FIG. 5. Tungsten probes. a. Spacing between quartz tube and probe accomplished by distance wire. b. Spacing made by splitting the tungsten wire. c. Spherical probe.

thin tungsten wire, insulated by a pile of small disks of glass cloth which damped the mechanical oscillations. The probe was moved by tilting the whole tube. The cathode in this tube was later changed from a Ba-Sr-oxide cathode to a Ba cathode, then to a Phillips L-Cathode. Tube No. 3 was also a demountable tube, with a silver-solder joint instead of the lead joint, so that it could be baked out at a temperature up to 430°C. It had a coaxial electrode arrangement (anode diameter 8 cm, Fig. 4). The anode cylinder was made of (nickel-iron) wire mesh so that the position of the probe and the discharge could be closely observed. The probe could be moved to any point between anode and cathode. The magnetically operated screw arrangement provided the movement along an axis which could be inclined by means of a pliable bellows. The heavy cathode and heater leads made it possible to use this tube for measurements with an indirectly heated Ta cathode which was able to give emission currents up to 1 amp (heater current 60 amp). The tungsten heater was of bifilar arrangement to reduce magnetic influences on the discharge. It was completely shielded by the cathode so that no electrons emitted from the heater could penetrate into the discharge proper (heater voltage 6v). The vacuum system was arranged in the following way: mercury diffusion pump, a ball joint mercury shut-off, stop cock trap inlet for xenon, argon, and helium, dry ice cooling trap, Pirani- and ion-gauge, test tube.

The Probes

The probe and its insulator had to be extremely small in size because the "ball of fire," which is characteristic for these low voltage discharges, in most cases shuns the probe.[11] The plasma density was normally so high (between 10^{10} and 10^{13} per cm^3) that it was possible to use short probes without appreciably influencing the collecting area of the probe by the ion sheath around the probe insulator.[12] Pictures of typical probes used here are shown in Fig. 5. Between the quartz insulator

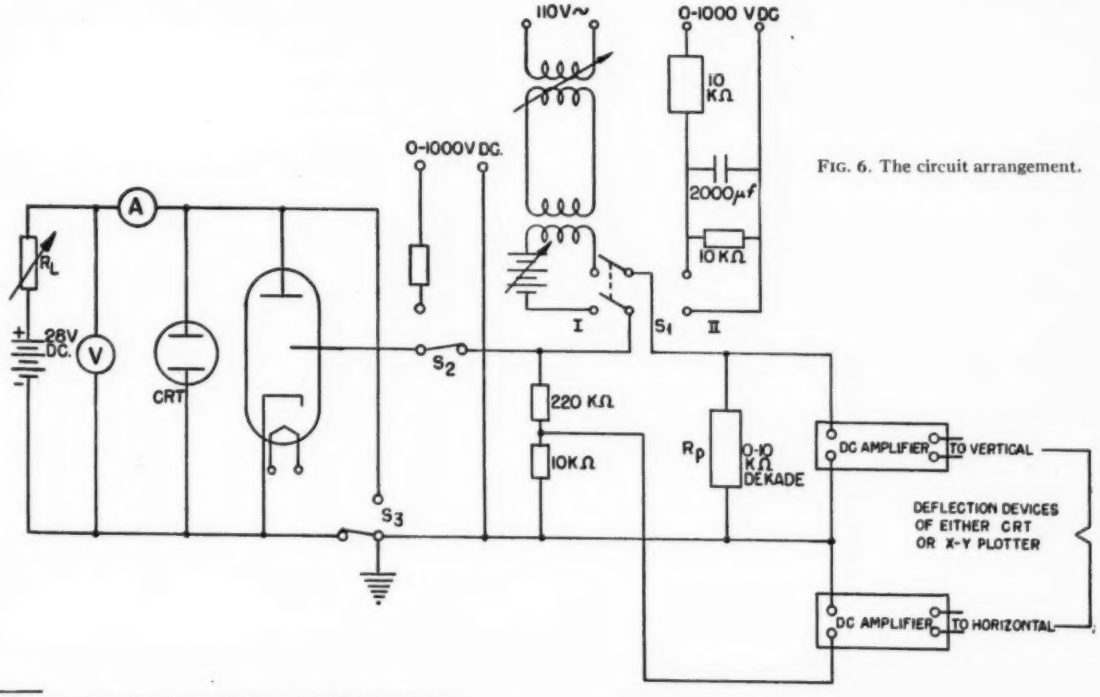


FIG. 6. The circuit arrangement.

[11] W. Funk and R. Seeliger, Z. Physik **113**, 203 (1939).
[12] R. H. Sloane and K. G. Emeleus, Phys. Rev. **44**, 333 (1933).

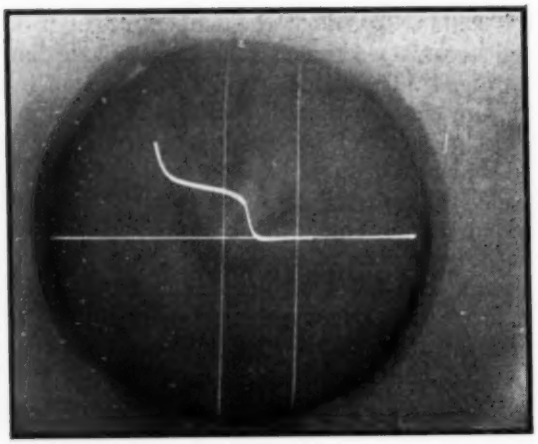


FIG. 7. Normal probe characteristic.

tube and the probe, a small spacing was provided to avoid conducting connections between the probe and the outer surface of the quartz tube. The mechanical connection between quartz tube and probe wire cannot be made too far from the probe end. Otherwise, the different thermal expansion coefficients of quartz and tungsten cause the probe to project more out of the quartz tube at higher temperatures. The spherical probes were made by melting the end of a tungsten wire in helium with an electric arc.

The Circuits

The probe circuit and the arrangement for recording the probe characteristic are shown in Fig. 6. R_p is a resistor decade in series with the probe to give voltages proportional to the probe current, provided all parts on the probe side of this resistor are so insulated from ground that leakage currents are small compared with the probe currents. This voltage and the probe voltage itself are connected through separate dc amplifiers to the horizontal and vertical deflection devices of either an oscillograph or an X-Y plotter. With switch S_1 in position I, the circuit is closed for oscillographic 60-cycle display and in position II for plotter recording. The X-Y plotter writes the probe characteristic with a pen on a graph paper of maximal 30×30 in.^{2}. At a writing speed of about 8 in. per second, the dynamic accuracy is claimed to average 0.05 percent. In order to smoothly regulate the applied voltage between zero and about 20 v at a desired velocity, a resistor-condensor circuit connected to a controllable 1000-v power supply, as shown in Fig. 6, was very helpful. Switch S_2 connects the probe to another power supply for cleaning purposes. The probe could be made positive and heated up to 2500°C as an anode, or it could be made negative for sputtering by ion bombardment. When the anode has settled down to a constant work function, it is better to use the anode instead of the cathode as reference electrode (Switch S_3), because plasma potential changes mostly occur near the cathode. Fluctuations in the cathode heating power had to be kept down as much as possible. This was done with a constant-voltage transformer.

An oscilloscope between anode and cathode indicated if the discharge was free of oscillations. This precaution is absolutely necessary for getting reliable measurements. Oscillations of the kind described by different authors are characteristic for these low voltage discharges, especially when oxide cathodes are used.[11,13,14] They can, however, be avoided easily by raising the discharge current above a certain limit.

EXPERIMENTAL PROCEDURES AND RESULTS

Oscillographic Observations

The Normal Probe Characteristic

Figure 7 shows a normal probe characteristic with the 60-cycle probe voltage in the horizontal direction and the probe current in the vertical direction. The vertical line on the right side indicates cathode potential; the left one is recorded at 10 v positive to the cathode. Electron currents to the probe raise the curve above the horizontal axis, whereas ion currents do the opposite. In the case of Fig. 7, the complete characteristic is on the screen, so that the small ion currents cannot be discerned. Starting from the right side, the probe first collects ions, then, at about +5v (reference: cathode), electrons begin to reach the probe; the "knee" at about +7v indicates the plasma potential. Above plasma potential all available electrons around the probe are collected, and with increased probe voltage a more or less saturated part results, depending on the ratio of probe diameter to electron sheath thickness. When the difference between probe potential and plasma potential reaches the lowest excitation potential of the gas, another rise in electron current appears, due to an increase in plasma density around the probe caused

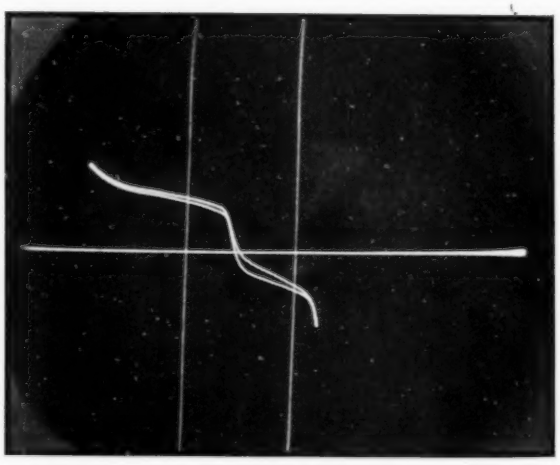


FIG. 8. Electron emitting probe.

[13] H. Kniepkamp, Z. techn. Physik **17**, 398 (1936).
[14] J. D. Cobine and C. J. Gallagher, J. Franklin Inst. **243**, 41 (1947).

by the ionization of the gas by the incoming electrons. Up to the plasma potential knee the probe currents normally stay below a few percent of the discharge current. In the ionizing region, however, the probe currents may rise to the order of the discharge current. Due to the high current density and increased electron velocity, the probe then becomes white hot, but it cools off considerably during the rest of the cycle. This automatic heating can be adjusted so that the probe stays clean. In the case of the heated probe, electron emission may cause spurious results; as long as the W-probe was clean and its temperature stayed below 2000°C, its emission was negligibly small.[15]

Emitting Probe

Figure 8 shows how electron emission from the probe affects the probe characteristic. In this case, the probe was located near the cathode for a while so that emitting material was evaporated to the probe surface. This process could be intensified by flashing the discharge; that means shorting the load resistor of the discharge for a moment. After removing the probe from the vicinity of the cathode, it was heated during the positive cycle (left side) to red heat (at too high a temperature the Ba-deposit evaporated too quickly); it stayed hot enough to emit electrons as soon as it became negative against plasma. Emitted electrons, of course, appear in the same direction as collected ions. This emission current is superimposed to the collected plasma electron and ion currents. This causes a sharper knee at plasma potential and a departure below the horizontal axis, again with a more or less saturated part. When the emitted electrons are accelerated to a velocity which is higher than the lowest excitation potential of the gas, they start to ionize and cause a rise in negative current similar to that of the positive one on the left side. The hysteresis which appears *above* the plasma potential (left side), where no emission can take place, is probably caused by the difference in probe temperature in the up and down branches, (the lower branch belongs to the hotter probe) which affects the gas and plasma density around the probe. The hysteresis which appears *below* plasma potential is caused by the changing electron emission while the probe cools off.

Incoming as well as outgoing electrons are accelerated to ionizing energies in the fields existing in the corresponding electron or ion layers. The oscillogram, however, does not show the potential difference equivalent to the accelerating layers, but the work function is either added (incoming) or subtracted (outgoing electrons). Therefore, in spite of some intrinsic asymmetry between the ionization effects of the incoming and the outgoing electrons, such an

[15] M. L. E. Oliphant, Proc. Roy. Soc. (London) **ACXXXII**, 631 (1931).

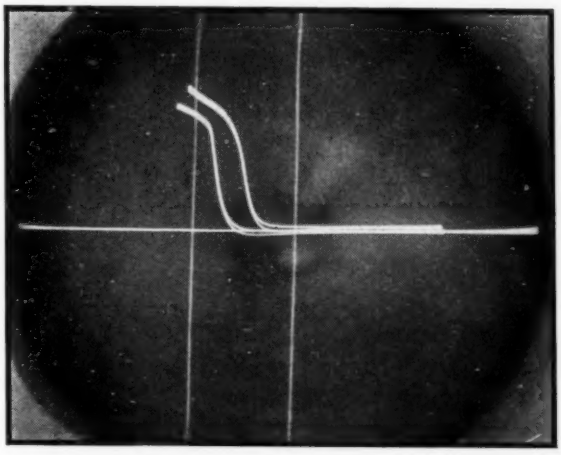


FIG. 9. Influence of the work function on the probe characteristic. Hysteresis is caused by shortcomings in the circuitry.

influence of the work function of the probe is clearly visible in Fig. 8. Here the knee is shifted to the right, from the middle between the regions of beginning ionization at either end of the curve, by an amount roughly equal to the work function of tungsten. Contrary to the *emitted* electrons, many of the *incoming* electrons are collected by the probe before they have a chance to excite or ionize gas atoms. This causes the upper ionization knee to be much flatter and more rounded than the lower one.

Influence of the Probe Work Function

Figure 9 is made with the probe at the same place, under the same discharge conditions with a different work function. The left curve was drawn with the probe thoroughly cleaned by making it white hot before the measurement; the right one was drawn with the probe covered with low work function material. The potential difference between the two curves is about 2.5 v.

In connection with the contamination of the probe surface with Ba compounds, a peculiar effect could be observed when the probe, located near the cathode, was connected to a positive voltage of about 15 v, with a low ohmic resistor in series. The potential between the clean tungsten probe and the plasma at the start is not high enough for the incoming electrons to ionize the gas, and the Ba compounds settle on the probe surface. The lowering of the probe work function causes the accelerating voltage for the plasma electrons to increase; the gas near the probe starts to become ionized, the probe current rises, the probe is heated, and the Ba compounds finally evaporate. This brings the probe back to its original clean state. The frequency of this "light house" effect is in the order of seconds and decreases, as expected, with the distance from the cathode.

Probe Studies with an X-Y Plotter

To study and evaluate probe characteristics, especially for changing work functions by taking photographic pictures from a scope, is somewhat cumbersome and not too exact. An *X-Y* plotter, however, turned out to be an ideal device for this purpose; in the following studies, the oscillographic method is only used to get a first over-all picture of the probe characteristic before using the plotter.

Poorly Conducting Probe Surface Layers

Curves 1 and 2 in Fig. 10 show a case similar to that of Fig. 9 but recorded with the plotter. Curve 1 belongs to the clean *W* probe and curve 2 to the same probe covered with Ba oxide. The two curves not only differ with respect to the influence of the work function of the probe, which should result in a parallel shift along the abscissa, as was described above, but also seemingly show a different velocity distribution of the plasma electrons, having different slopes. On the other hand, when Ba is used instead of BaO for lowering the work function of the probe, a perfectly parallel shift of the curves along the abscissa was obtained without any change in the slope. Curve 3 shows the effect of an outside resistor in series with the clean probe (between probe and S_2 in Fig. 6). It yields the same general curve form as appeared with BaO on the probe. From these observations it is concluded that the resistance of the BaO layer is involved because the recorded probe voltage is not identical to the actual probe surface potential (disregarding the change in work function). The difference obviously is caused by the iR drop through the insulating layer and is a function of the probe current.

Similar observations can be made in any discharge with a Ta probe if it is covered with a thin layer of Ta oxide which has a low conductivity and a high evaporation temperature. This layer could not be removed by heating the probe to white incandescence. However, it could easily be eliminated by cathode sputtering (100 v negative for about 1 minute was sufficient).

Such flattened probe characteristics, with poorly developed plasma potential knee originating from the resistance of surface layers, are very common, usually very stable and reproducible. Therefore, one must be quite careful in making sure that the probe is perfectly clean before drawing reliable conclusions concerning electron velocity distribution and plasma potentials.

Studying the deviations in probe characteristics of covered probes from those of clean metal probes in the same position might make it possible to measure conductivities of surface layers. For instance, it would be possible to study Ba-Sr oxide layers under different temperatures, or activation conditions, and up to high current densities. This would avoid the use of an additional electrode (which is represented here by the plasma).

Different Work Function at Different Parts of a Probe

Another case of flattened probe characteristics arises when different parts of the probe have different work functions. The low work function sections determine the lower part of the electron current curve because they start to collect electrons sooner than the high work function sections. The plasma potential knee, however, is determined by the part of the probe surface with the highest work function because only above this potential are all electrons around the probe collected. Hence, a flattened intermediate curve between low and high work function probe characteristic results with the same consequences concerining the reliability of plasma electron temperature measurements as in the case of low conducting surface layers.

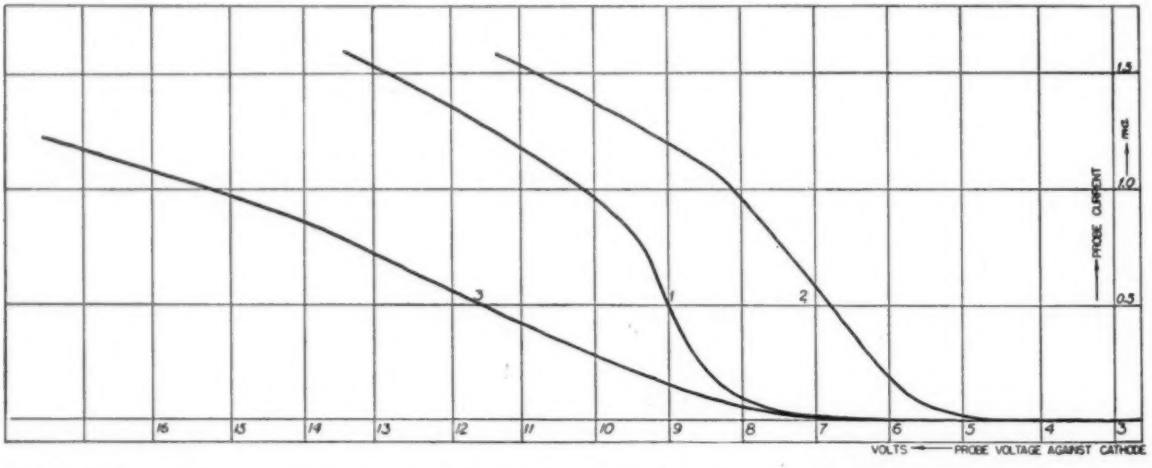


FIG. 10. Influence of a poorly conducting surface layer. Tube No. 2. Discharge current $I_a = 1$ amp. Voltage drop $V_a = 5$ v. Xenon 200μ. Probe halfway between cathode and anode. Curve 1: Clean probe. Curve 2: Probe covered with Ba oxide. Curve 3: Outside resistor in line with the clean probe.

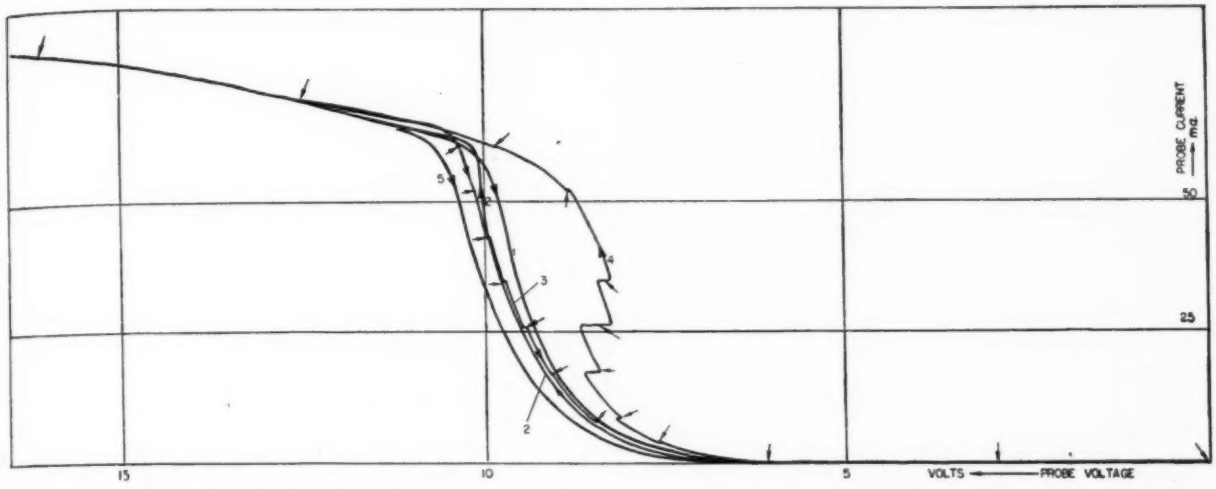


FIG. 11. Probe characteristic changing with time. Tube No. 1 with oxide cathode. $I_a=1.2$ amp, $V_a=5.85$ v. Xe 150μ. Probe a (Fig. 5) 10 mm away from cathode. Curve 1: Clean probe recorded downward. Curve 2: Immediately thereafter upward. Curve 3: Downward with 30 sec waiting periods at the indicated points (arrows). Curve 4: The same upward. Curve 5: After keeping probe for 15 minutes near plasma potential.

Work Function of a Probe Changing with Time, Probe Current, and Probe Voltage

General view.—Figure 11 gives an example of how intricate probe measurements can be in oxide cathode tubes under unfavorable conditions when gradual changes of time come into play. Curve 1 was recorded downward (toward cathode potential) immediately after the probe was thoroughly cleaned by raising its temperature to about 2000°C and curve 2 immediately thereafter upward. Curve 3 was taken downward after cleaning the probe but with waiting periods of 30 seconds at the indicated points (arrows) and curve 4 the same way, upward. Curve 5 belongs to a probe which was kept at plasma potential for 15 minutes and was recorded downward. In order to check if the discharge had changed during all these measurements, curve 1 was repeated and found to be unaltered.

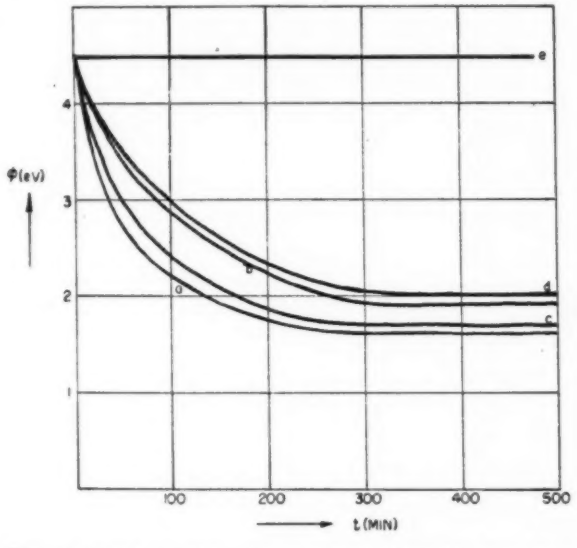


FIG. 12. Work function of probe decreasing with time due to deposition of barium. $I_a=0.5$ amp, $V_a=4.5$ v. Xe 150μ. Tube No. 2 with Ba cathode. Probe 5 mm away from anode. Curve *a*: Discharge switched off. Curve *b*: Discharge running and probe at cathode potential. Curve *c*: Probe at +3 v. Curve *d*: Probe at +5 v. Curve *e*: Probe at plasma potential between the measuring procedures.

In order to study and to separate the different influences which may change the probe characteristic with time, it was necessary to start with less complicated discharge conditions and, thereafter, build toward the more complex ones.

Tube with tantalum cathode.—The simplest case would be a low voltage drop tube (small sputtering effects), which, of course, must be free of gas impurities and oscillations and has only pure metal electrodes (including cathode) of the same material as the probe. Such a tube with Ta cathode is under construction and will be very helpful in studying such minute effects as work function changes of a pure metal with temperature, or changes of surface conditions as might be caused by adsorbed noble gases, or by sputtering by ion bombardment. Preliminary tests (tube No. 3, Ta cathode) showed, furthermore, that the investigation of the work function change of a probe provides a sensitive methode for measuring the rate of sputtering. For this purpose another electrode, with different work function, is cathode-sputtered so that the probe is covered with a deposit of material with a different work function. This method is so sensitive that even measurements at very low ion velocities can be performed in a comparatively short time. This provides a possibility to determine if an ion velocity threshold for different materials [16,17] exists.

Tube with barium cathode.—In a Ba cathode tube (Tube No. 2), Ba was evaporated from a Ba core-nickel-wire onto a Ni base. Here the work function is

[16] K. H. Kingdon and I. Langmuir, Phys. Rev. **22**, 148 (1923).
[17] A. W. Hull, Trans. Am. Inst. Elec. Engrs. **47**, 755 (1928).

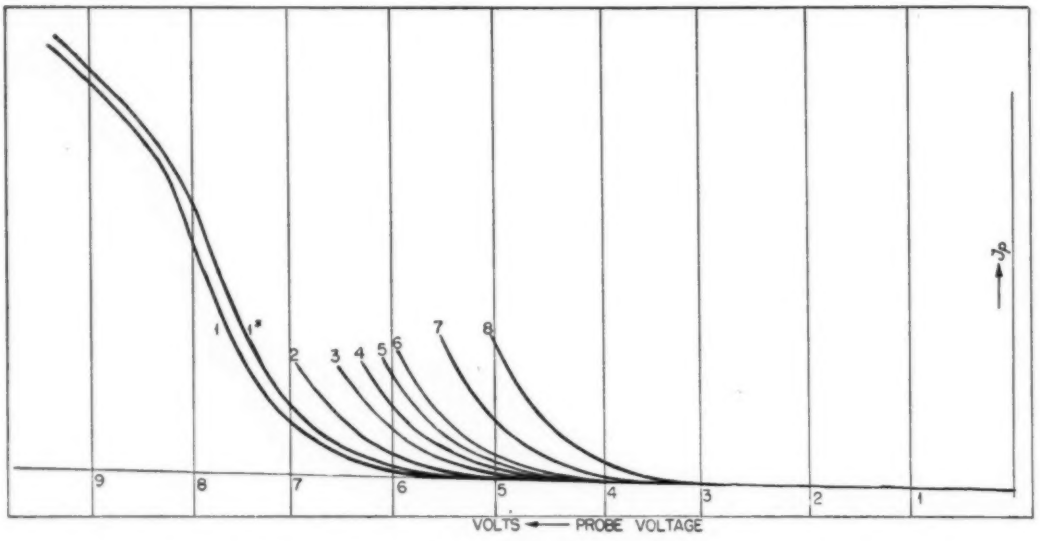


FIG. 13. Work function change caused by barium deposition. Data as in Fig. 12. Probe kept at cathode potential between recordings. Curve 1: Clean probe. Curve 1*: Sputtered probe. Curves 2–6: 20 minute intervals. Curve 7: After 60-minute interval. Curve 8: Final characteristic.

being lowered by the pure Ba alone, without complications of poor conducting surface layers.

Figure 12 "*a*" shows how the work function of the clean tungsten probe (work function assumed to be 4.50 v) decreases as a function of time when the discharge is switched off between the measuring points. During the measuring operation itself (which takes some seconds), the probe current must be kept below a value which would cause the Ba to re-evaporate or to migrate over the surface. When the discharge is kept going this curve and the final value of the work function depend on the potential at which the probe was kept during the intervals between the measurements, because this potential determines how much of the Ba which reaches the probe is sputtered or re-evaporated. Figure 13 shows the shifting of the probe characteristic when the probe was kept at cathode potential. Curve 1 belongs to the clean tungsten probe, curves 2–6 to 20 minute intervals of sputtering, curve 7 is recorded after another 60 minutes, and curve 8 is the final one. The resulting work function change *vs* time is plotted in Fig. 12 curve "*b*" and shows that at cathode potential some sputtering takes place and less Ba is deposited. The evaporation of Ba from the probe caused by increased probe temperature in consequence of ion bombardment in this case is negligibly small, because the ion currents are less than 1 ma and the ion velocities less than 6 v. This results in a heating power which is at least one order of magnitude less than found necessary (with collected electron current) for getting a noticable Ba evaporation, even if the neutralization energy is added.[15,18] Curve *c* shows the result when the probe was kept at 3 v positive against cathode and indicates that the sputtering influence decreased. At 5 v positive to cathode (curve *d*) the increased probe temperature (due to the collected electron current) causes some Ba to migrate or to re-evaporate from the probe. At plasma potential (curve *e*) the probe temperature is sufficient to prevent the deposition of Ba and the probe stays clean.

Figure 14 shows that, with the probe near the anode, the rate of Ba deposition on the probe surface is much higher without a discharge. Curve 1 belongs to the clean tungsten surface; curve 2 is recorded after the discharge had been going for 5 minutes (the probe potential was kept at 3 v negative against plasma potential; this gave the smallest sputtering and evaporation effects), and curve 3 the same after another 5

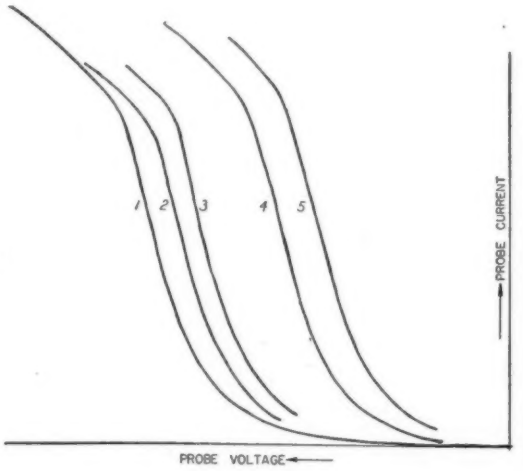


FIG. 14. Rate of barium deposition with and without running discharge. Tube No. 2 with Ba cathode. I_a=0.5 amp, V_a=4 v. Xe 150μ. Curve 1: Clean probe. Curves 2 and 3: After 5-minute intervals while the discharge was running. Curves 4 and 5: After 5-minute intervals while the discharge was switched off.

[18] W. Moltham, Z. Physik **98**, 227 (1935).

minutes. Curves 4 and 5 were made with the same time intervals with the discharge switched off between the measurements. With the probe near the cathode, the results are just reversed, that is, the higher rate of deposition was observed when the discharge was kept running. One possible explanation is that a high percentage of the Ba becomes ionized and returns back to the cathode under the influence of the electrical field. The Ba vapor may become ionized to a high degree, owing to the low excitation and ionization potentials of Ba together with the comparatively high electron temperature of the plasma electrons in these low voltage discharges. The returning of the Ba back to the cathode could be directly checked by measuring the lifetime of a nickel base Ba cathode. With the discharge running this lifetime is much longer than without the discharge. Especially in the region near the cathode the rate of deposition on the probe may be so high and consequently the work function may change so rapidly that probe measurements can be seriously affected.

When a Phillips L cathode was used under equivalent discharge conditions, signifying high cathode temperature and consequently high Ba evaporation, it was impossible to get reliable results with a static measuring method. The only way to avoid the heavy Ba deposition on the probe was to keep the probe hot enough at all times, for instance, by using a heated probe or the oscillographic method with automatic probe heating during the positive part of the cycle. Another possible way for making static probe measurements is to wait until the probe work function has reached a final low value. This would be applicable only in discharges or regions of a discharge with low plasma density where evaporation or sputtering from the probe is negligibly small.

Figure 15 shows the effect of a probe which was not properly fixed in the quartz tubing (the mechanical connection between probe and quartz tubing was too far away from the probe end. Curve 1 was recorded downward, curve 2 immediately afterward, the opposite way. In curve 1 a longer portion of the probe projects from the quartz tubing due to the higher probe temperature and the different thermal expansion coefficients and indicates, therefore, a faulty higher density than in curve 2.

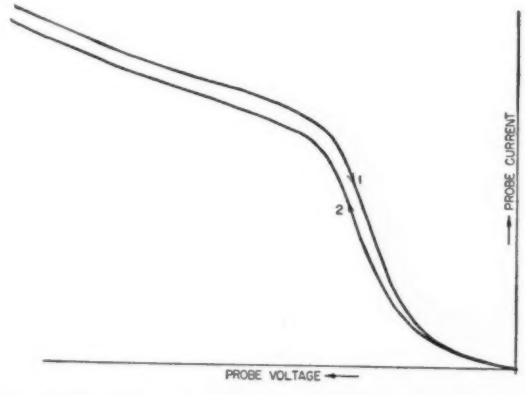


FIG. 15. Change of the effective probe area in a faulty constructed probe with temperature. Curve 1: Recorded downward with the probe hot at the start. Curve 2: Recorded upward with the probe cool at the start.

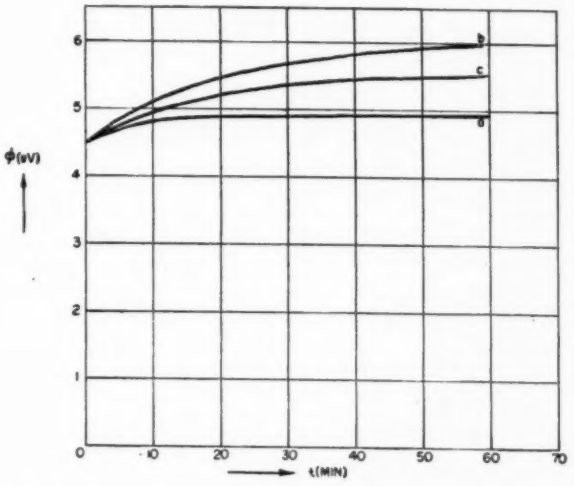


FIG. 16. Increasing work function in an oxide cathode tube. Probe (*b*) 5 mm away from anode. Tube No. 2 $I_a=1.2$ amp. $V_a=5.85$ v. Xe 150μ. Curve *a*: At cathode potential. Curve *b*: At 1 v negative against plasma potential. Curve *c*: At 5 v positive against plasma potential between the measuring procedures.

Cleaning the probe by ion bombardment resulted in a work function about 0.1 v lower than for the heated probe (curve 1* in Fig. 13). The amount of this work function difference seems to be about the same for different velocities of the bombarding ions and for different periods of bombardment. The work function goes back to the "heated" value when the probe temperature is raised to a dull red.

Tube with oxide cathode.—Much more complex phenomena are encountered when oxide cathodes are used because of the oxygen liberated from these cathodes.[19] The oxygen may not only be adsorbed or chemically bound to the probe but will form insulating BaO layers as well.

Comparatively simple conditions again prevail in oxide cathode tubes when the probe is far enough away from the cathode. In this case only minute amounts of Ba or BaO reach the probe because most of the Ba (and compounds) return to the cathode, as indicated. The predominating influences on the work function are in this case the adsorption of oxygen[20] and the evaporation and sputtering of the oxygen away from the probe. Curve *a* in Fig. 16 shows that, due to the sputtering of the probe surface, (the probe potential between the measurements was kept at cathode potential, in this case 8.5 v negative against plasma potential) the work function increase by absorption of oxygen is quite small. With increasing probe

[19] R. H. Sloane and C. S. Watt, Proc. Phys. Soc. (London) **LXI**, 217 (1948).

[20] J. A. Becker, Trans. Faraday Soc. **28**, 148 (1932).

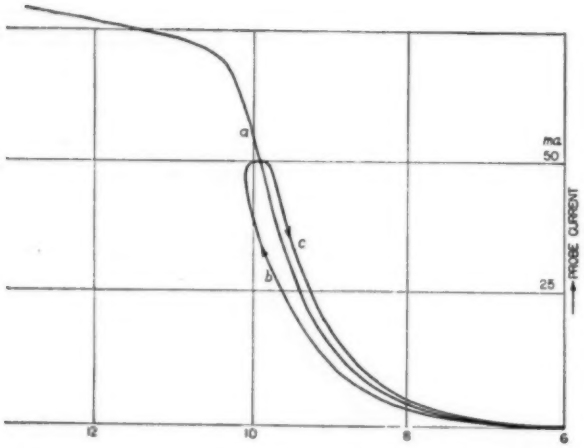


FIG. 17. Probe near an oxide cathode. Tube No. 2. Probe *b*, 10 mm away from cathode. $I_a=1.2$ amp. $V_a=5.85$ v. Xe 150μ. Curve *a*: Clean probe. Curve *b*: Recorded upward after the probe was kept at 6 v negative against plasma potential for 1 minute. Curve *c*: Downward after the probe was kept at 50 ma probe current for 1 minute.

potential, the amount of sputtered oxygen decreases, and the work function increase reaches a maximum value (curve *b* at plasma potential) before the evaporation of oxygen, due to the higher probe temperature, reduces it again (curve *c* at 5 v positive against plasma potential). The probe temperature at which the oxygen layer begins to decay is much higher than the corresponding temperature for the Ba layer, as is to be expected from the higher binding energies of oxygen. The adsorption of oxygen on the probe takes place with the probe potential either positive or negative with respect to the plasma potential, which shows that the oxygen is not ionized.

When the probe is nearer to the cathode, the deposition of Ba and BaO takes place in addition to that of oxygen. The result in this case is that the work function of clean tungsten first increases, then reaches a maximum, and finally decreases below that of pure tungsten. This is shown in Fig. 17 where curve *a* is the characteristic of the clean tungsten probe. After the probe potential was kept at 6 v negative against plasma potential for 1 minute, the characteristic was recorded upward in curve *b* and shows the increased work function. After waiting at 50-ma probe current for 1 minute, the work function goes down to curve *c*. The explanation for this might be the following: Ba and O_2 deposited separately in very small amounts form a microscopic surface pattern with discrete spots of higher and lower work function than that of the tungsten base. At higher probe temperatures the Ba atoms start to migrate over the surface and to combine with the O_2 to Ba oxides, thereby, decreasing the overall work function. The speed of this "activation" is a function of the probe temperature; therefore, when the probe potential is raised above the plasma potential, the work function decreases so fast that it is no longer possible to follow this change with the plotter. When the temperature of the probe is raised still further, it causes the deposited compounds to evaporate more and more.

Less open to interpretation is the case of a probe which was covered with oxide from the cathode by "flashing" the discharge. Curve *a* in Fig. 18 shows the

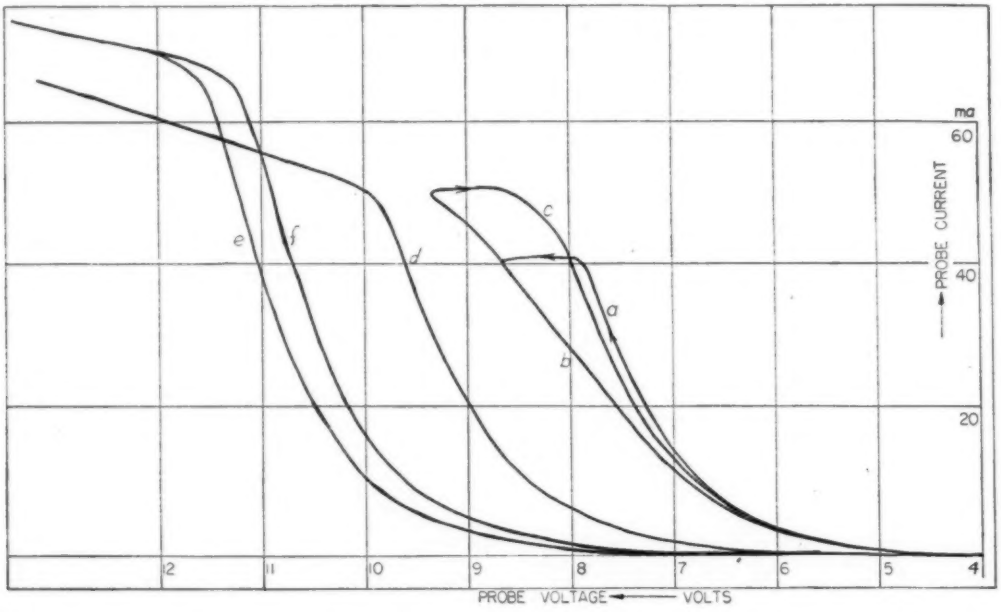


FIG. 18. Probe covered with Ba compounds. Data as in Fig. 17. Curve *a*: Covered with Ba compounds. Curve *b*: After keeping the probe for 2 minutes at 40 ma probe current. Curve *c*: After keeping the probe for 5 minutes at 50 ma probe current. Curve *d*: After raising the probe temperature to about 900°C. Curve *e*: After raising the probe temperature to about 1300°C. Curve *f*: After raising the probe temperature to about 1800°C (clean probe). The indicated plasma density changes are produced by fortuitous changes of discharge current.

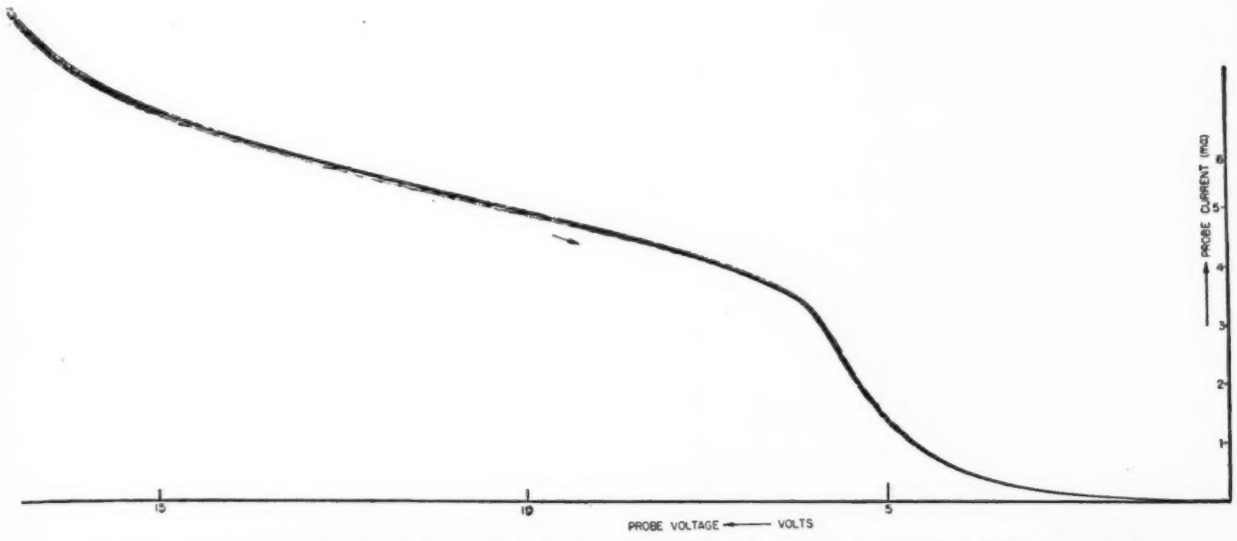


FIG. 19. Reproducibility of a probe characteristic. Tube No. 3. Ta-cathode. $I_a = 0.4$ amp, $V_a = 1.75$ v. Xe 150μ. Probe characteristic recorded 10 times with 1-minute intervals.

characteristic of a probe after "flashing" with no signs of resistance and the lowest work function observed in the case of an oxide cathode. Raising the probe temperature by drawing an electron current of 40 ma resulted in the formation of a resistance layer as shown in curve *b*. Raising the probe temperature further destroyed this resistance layer, but still almost no change in the work function is seen (curve *c*). Raising the probe current still further resulted in a higher work function with no more signs of resistant properties as recorded in curve *d*. This was plotted after the probe temperature had been raised to about 900°C. Raising

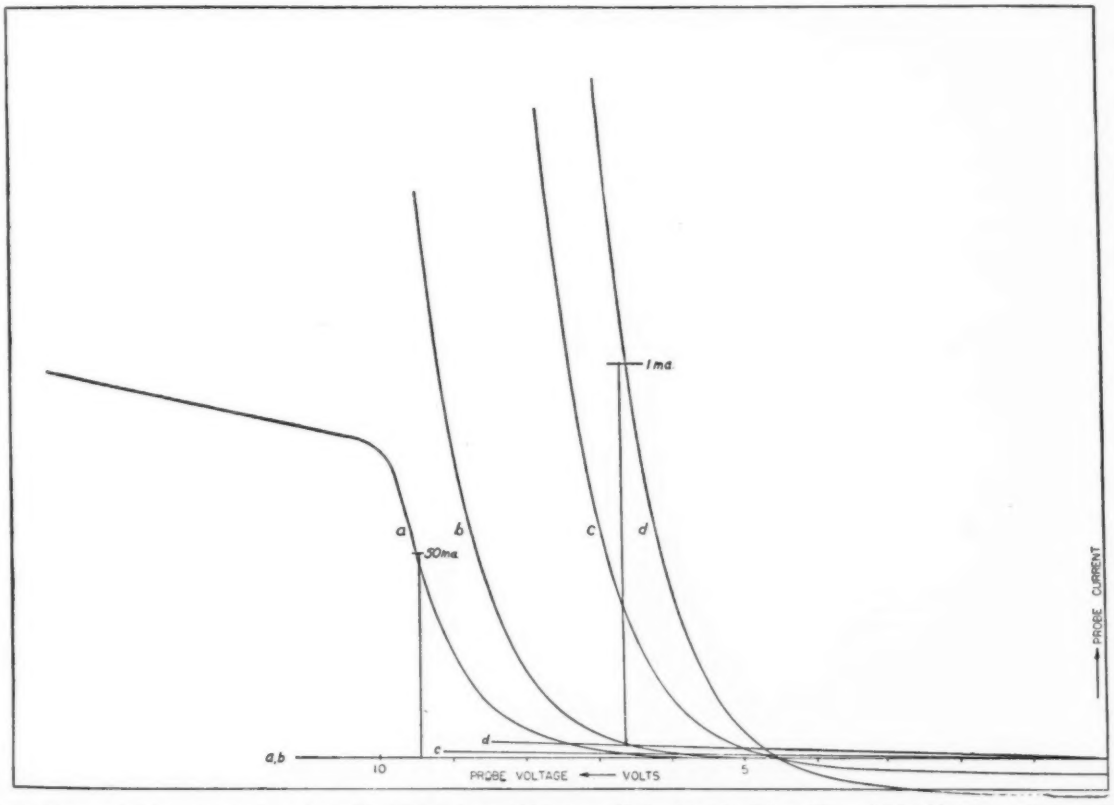


FIG. 20. Characteristic of a clean probe with different amplifications. Tube No. 2 with oxide cathode. Probe *c*, 10 mm away from cathode. $I_a = 1.2$ amp, $V_a = 5.85$ v. Xe 150μ. Curve *a*: Complete characteristic. Curve *b*: Amplification 3.2 (zero current axis the same). Curve *c*: Amplification 32 (axis *c*). Curve *d*: Amplification 94 (axis *d*).

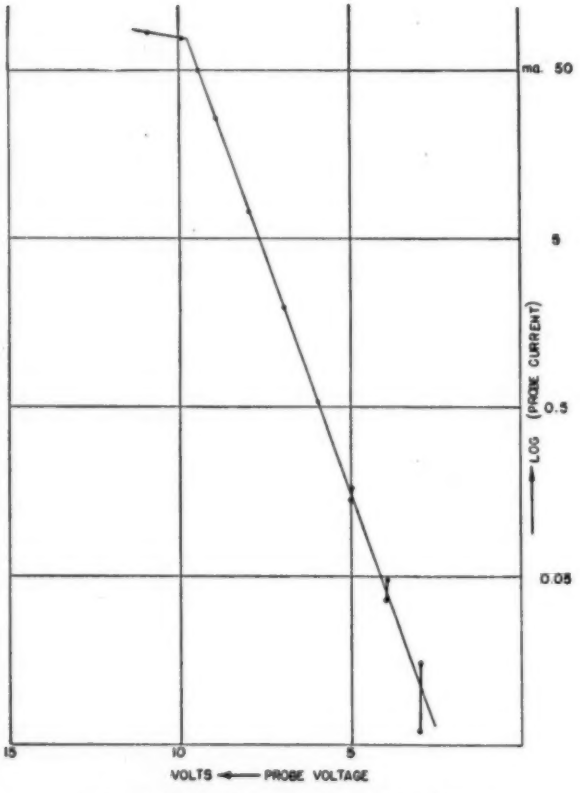


FIG. 21. Evaluation of Fig. 21. Log of probe current *vs* probe voltage.

the temperature to 1200°C yielded a work function higher than that of the clean tungsten (curve *e*). Finally, after heating the probe to 1800°C, the characteristic returned to that of the clean tungsten.

From Fig. 19 the reproducibility of a probe characteristic may be seen. The probe was properly cleaned immediately before recording the characteristic, the discharge being held constant. The characteristic was recorded 10 times with one-minute intervals and the picture shows the actual photograph of the plotting paper (in all other plotter pictures, the curves were retraced for the sake of clarity).

In order to investigate the plasma electron velocity distribution, the current scale had to be increased by recording the lower part of the probe characteristic with a higher amplification. Figure 20 shows that the zero current axes are not the same for higher amplifications due to a small leakage current. No deviation from the straight line in the plot of log (probe current) *vs* probe voltage (Fig. 21) could be found, although the probe in this experiment was only 10 mm away from the cathode (at $150\mu Xe$), and almost 4 orders of magnitude in probe current are covered. The limits of the possible error which arise at small electron currents, due to the shifting of the zero current line by collected ion current, are indicated in the picture. This result is in contradiction to the deviations from the Maxwellian velocity distribution found by Druyvesteyn and others in the same type of discharge (oxide cathode, low voltage diode). The facts, that these deviations did not arise in a tungsten cathode tube[2] and that no mention is made in these early publications concerning the probe cleaning procedure, lends reasonable grounds for assuming that the probe had changed its work function during the measuring procedure. This suspicion gains support through the trend of the reported deviations from the Maxwellian distribution. The work function would have been lower at smaller collected electron currents than at higher. This is in agreement with our results as soon as Ba compounds are involved and evaporate from the probe at higher probe temperatures.

ACKNOWLEDGMENTS

The authors wish to express their thanks to their associates in the Components and Systems Laboratory, Electron Tube Branch, who gave help in many respects, especially to Otmar Stuetzer and Ludwig Mayer for stimulating discussions, to Amos H. Dicke as supervisor for his interest in this work, and to Emil Benz for his excellent job in glass work.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 9 SEPTEMBER, 1952

On the Theory of Prediction of Nonstationary Stochastic Processes

R. C. DAVIS
U. S. Naval Ordnance Test Station, Inyokern, Pasadena Annex, California
(Received April 28, 1952)

We consider the following problem of prediction: During a finite time interval T the real valued function $S(t)+N(t)$ is observed, in which $S(t)$ is a signal and $N(t)$ is a linearly superimposed noise disturbance. The problem is to predict the value of a given linear functional of $S(t)$, the predictor formula having certain preassigned "optimum properties" among a certain class of predictors. In the case in which the mean value of $S(t)$ is known, the random components of $S(t)$ and $N(t)$ are strictly stationary, and the time interval T is infinite, a complete solution to this problem has been given by N. Wiener. (In the case of discrete time series, the solution was given by A. Kolmogoroff.) This theory has been extended by L. Zadeh and J. Ragazzini [J. Appl. Phys. **21**, 645 (1950)] to the case in which T is a finite time interval and the mean value of $S(t)$ is unknown but is restricted to be a polynomial in time. We extend the above theories to the case in which the random components of both $S(t)$ and $N(t)$ are nonstationary in time and merely possess finite continuous covariance and cross-covariance functions. The analytical tools of probability theory which we use are those developed independently by M. Loève and K. Karhunen in their studies of a class of stochastic processes usually termed processes of second order. Apparently these techniques are practically unknown outside of certain mathematical circles. In the opinion of the author these techniques are extremely powerful in the analysis of transient random phenomena in linear systems.

Finally we give an exposition of a method of prediction by the theory of conditional probabilities. This method is applicable when the form of the joint probability distribution of signal and noise is known and hence is applicable to many practical problems, since this distribution is often Gaussian. In this particular case the predictor formula given by the theory of conditional probabilities is identical with the usual linear predictor formula given by the theory of least squares. In case the signal is a Markoff process, the method of conditional probabilities yields a much simpler prediction formula than the usual method of least squares.

I. INTRODUCTION

LET $S(t)$ be a continuous random real valued time series (which we will denote as a signal) and let $N(t)$ be another continuous time series (which we denote by noise) linearly superimposed upon $S(t)$. We observe the sum $S(t)+N(t)$ during a finite time interval of duration T. For any random variable x we denote by Ex the mathematical expectation of the random variable x. This is defined in terms of the cumulative probability distribution of x, denoted by $F(a)$, so that

$$F(a)=Pr.\{x\leqslant a\}.$$

In terms of $F(a)$, Ex (if it exists) is given by

$$Ex=\int_{-\infty}^{\infty} x dF(x).$$

Having defined the mathematical expectation, we can now formulate the assumptions needed concerning the statistical nature of the fluctuations of $S(t)$ and $N(t)$.

Assumption 1. $ES(t)=m(t)$ *exists and is finite but unknown.*

Assumption 2. $EN(t)\equiv 0$.

Assumption 3. The following covariance functions exist, are finite, and are known:

$$E\{[S(s)-m(s)][S(t)-m(t)]\}$$
$$E\{[S(s)-m(s)][S(t)-m(t)+N(t)]\}$$
$$E\{[S(s)-m(s)+N(s)][S(t)-m(t)+N(t)]\}.$$

Moreover for every pair (s, t) *in the square* $0\leqslant s\leqslant T$, $0\leqslant t\leqslant T$, *the above covariance functions are continuous in the pair* (s, t).

Commenting on the above assumptions, in many applications of any theory of prediction the mean value $m(t)$ of the signal is unknown. Some *a priori* information is usually available concerning the nature of $m(t)$, and an attempt must be made to incorporate this information in a predictor formula. If sufficient previous observations are available, it may be possible to assume that $m(t)$ is well approximated by a polynominal of degree n. In other cases one may know from physical considerations that certain of the derivatives have known upper bounds. Assumption 2 is satisfied in most cases of interest. Assumption 3 is never satisfied, strictly speaking, since one never knows covariance functions exactly, but must obtain estimates from previous data or from theoretical considerations. It would be desirable to be able to develop a prediction theory in which one incorporates a specified degree of uncertainty regarding covariance functions. (In statistical language one would obtain confidence regions for the covariance functions by analyzing past data.) The theory is difficult enough to apply, however, even when one assumes the various covariance functions to be known exactly. The assumption of continuity of the covariance functions is essential to the theory and is satisfied in most engineering applications.

In order to avoid unnecessary complications in the analysis and to present the theory as clearly as possible, we limit our analysis of prediction theory in the sequel initially to the simple yet practical case in which we wish to predict $S(T+\Delta)$, $(\Delta>0)$, knowing the values of $S(t)+N(t)$ during a time interval $[0, T]$. We will indicate, however, the manner in which this simple case can be extended easily to the general case of prediction of any given linear functional of $S(t)$.

Before attempting to develop a prediction formula we require a result due to M. Loève[1] and K. Karhunen.[2]

[1] M. Loève, Compt. rend. **220**, 469 (1946).
[2] K. Karhunen, Ann. Acad. Sci. Fennicae, A I, 37 Helsinki (1947).

First we define a stochastic process $x(t)$ to be continuous in the mean during the time interval $[0, T]$ if

$$E[x(t+h)-x(t)]^2 \to 0 \text{ for every } t\epsilon[0, T] \text{ as } h\to 0.$$

Karhunen has shown (see reference 2) that continuity in the mean and continuity of the covariance function are equivalent concepts. If $x(t)$ is continuous in the mean with mean value function $m(t)$ and covariance function $r(s, t)$, he has shown that

$$x(t)=m(t)+\sum_{i=1}^{\infty} Z_i \frac{\varphi_i(t)}{(\lambda_i)^{\frac{1}{2}}}, \tag{1}$$

with convergence in the mean for every $t\epsilon[0, T]$. $\{\lambda_i\}$ and $\{\varphi_i(t)\}$ are the eigenvalues and the corresponding eigenfunctions of the integral equation

$$\varphi(s)=\lambda \int_0^T r(s, t)\varphi(t)dt,$$

and the $\{Z_i\}$ are random variables satisfying

$$\begin{aligned} EZ_i&=0,\\ EZ_iZ_j&=\delta_{ij}. \end{aligned} \tag{2}$$

Since

$$\int_0^T \int_0^T r(s, t)f(s)f(t)dsdt = E\left\{\int_0^T [x(t)-m(t)]f(t)dt\right\}^2 \geq 0,$$

we see that $r(s, t)$ is positive semidefinite in the square

$$0\leqslant s\leqslant T, \quad 0\leqslant t\leqslant T.$$

We now want to represent the information contained in a certain realization of $x(t)$ in a convenient manner. We form the Fourier coefficients $\{x_i\}$, $i=1, 2, \cdots$, of the realization with regard to a complete orthonormal system of functions defined over the time interval $[0, T]$. It is convenient to take $\{\varphi_i(t)\}$ as this system if it is complete; otherwise we make it complete by adding another orthonormal system to it in the usual way.

We now consider the general case in which the signal $S(t)$ and the noise $N(t)$ may be stochastically dependent. We introduce the three known covariance functions

$$\begin{aligned} r_S(s, t)&=E\{[S(s)-m(s)][S(t)-m(t)]\},\\ r_{SN}(s, t)&=E\{[S(s)-m(s)][S(t)-m(t)+N(t)]\},\\ R_{SN}(s, t)&=E\{[S(s)-m(s)+N(s)][S(t)-m(t)+N(t)]\}, \end{aligned}$$

which, by Assumption 3, are assumed to be known.

We have then

$$S(t)+N(t)=m(t)+\sum_{i=1}^{\infty} Z_i \frac{\varphi_i(t)}{(\lambda_i)^{\frac{1}{2}}}, \tag{3}$$

with convergence in the mean for every $t\epsilon[0, T]$. $\{\lambda_i\}$ and $\{\varphi_i(t)\}$ are the eigenvalues and corresponding eigenfunctions of the integral equation

$$\varphi(s)=\lambda \int_0^T R_{SN}(s, t)\varphi(t)dt. \tag{3a}$$

II. DETERMINATION OF THE OPTIMUM LINEAR PREDICTOR

We restrict ourselves naturally to the prediction of linear functionals of $S(t)$. If we restrict ourselves to continuous functions $S(t)$, F. Riesz[3] has shown that every linear functional of $S(t)$ can be written as

$$L(S)=\int_{-\infty}^{\infty} S(t)dg(t),$$

where $g(t)$ is a function of bounded variation in $(-\infty, \infty)$ and the choice of which is independent of $S(t)$. In order to demonstrate the techniques more clearly, we will consider now the simple yet important problem: we have observed $S(t)+N(t)$ during the time interval $[0, T]$ and we wish to obtain a prediction of $S(T+\Delta)$, $\Delta>0$, which possesses some optimum properties to be exhibited later. By a linear predictor we mean a functional

$$\int_0^T K(t)[S(t)+N(t)]dt,$$

where the weight function $K(t)$ is a function also of the particular point $T+\Delta$ at which prediction is desired. If the system of eigenfunctions $\{\varphi_i(t)\}$ in (3) is a complete orthonormal set, we can write

$$K(t)=\sum_{i=1}^{\infty} a_i\varphi_i(t) \qquad 0\leq t\leq T.$$

Hence if we introduce the Fourier coefficients $\{x_i\}$, where

$$x_i=\int_0^T [S(t)+N(t)]\varphi_i(t)dt, \tag{4}$$

we see immediately that we are restricting our class of predictors to the class

$$S^*(T+\Delta)=\sum_{i=1}^{\infty} a_ix_i. \tag{5}$$

We introduce now an additional assumption concerning the mean value function $m(t)$:

Assumption 4. $m(t)$ is a polynomial in t of degree n.

From Bonnet's form of the second mean value theorem for integrals, we have

$$\int_0^T t^j\varphi_i(t)dt=T^j\int_{\xi_j}^T \varphi_i(t)dt, \quad 0\leq\xi_j<T.$$

[3] F. Riesz, Compt. rend. 149, 974 (1909).

Let

$$M_i=\underset{j=0,1,2,\cdots,n}{\text{Maximum}}\left|\int_{\xi_j}^{T}\varphi_i(t)dt\right|.$$

We now make the following assumption:

Assumption 5. $\sum_{i=1}^{\infty}\lambda_i M_i^2<\infty$.

The condition in Assumption 5 is an interesting generalization of a condition due to Grenander[4] for the existence of an unbiased minimum variance estimate for the mean value of a continuous stochastic process when the mean is a constant β_0. In this particular case, our M_i above becomes

$$M_i=\left|\int_0^T\varphi_i(t)dt\right|.$$

Since we are assuming a more complicated mean value function, namely, a polynomial of degree n, it is natural that we require a somewhat stronger condition.

We now state and prove the following:

Theorem. Among the class of all linear predictors of $S(T+\Delta)$, there exists a unique unbiased predictor possessing minimum expected squared error.

Proof: By definition $S^*(T+\Delta)$ is unbiased if

$$ES^*(T+\Delta)=m(T+\Delta). \tag{6}$$

Clearly the error in any predictor is

$$S^*(T+\Delta)-S(T+\Delta),$$

and the expected squared error is

$$\begin{aligned}J&=E\{S^*(T+\Delta)-S(T+\Delta)\}^2\\&=E\{S^*(T+\Delta)-m(T+\Delta)+m(T+\Delta)-S(T+\Delta)\}^2\\&=r_S(T+\Delta,T+\Delta)+E[S^*(T+\Delta)-m(T+\Delta)]^2\\&\quad+2E\{[S^*(T+\Delta)-m(T+\Delta)][m(T+\Delta)-S(T+\Delta)]\}.\end{aligned}$$

Now from Eqs. (3) and (4) we see that

$$x_i=(m,\varphi_i)+\frac{Z_i}{(\lambda_i)^{\frac{1}{2}}}. \tag{7}$$

Since the $\{Z_i\}$ are stochastically uncorrelated we see from Eqs. (5) and (7) that the condition that the predictor $S^*(T+\Delta)$ be unbiased (Eq. (6)) becomes

$$\sum_{i=1}^{\infty}a_i(m,\varphi_i)=m(T+\Delta). \tag{8}$$

Hence,

$$S^*(T+\Delta)-m(T+\Delta)=\sum_{i=1}^{\infty}\frac{a_i}{(\lambda_i)^{\frac{1}{2}}}Z_i$$

and

$$E[S^*(T+\Delta)-m(T+\Delta)]^2=\sum_{i=1}^{\infty}\frac{a_i^2}{\lambda_i}.$$

Considering the last term in J,

$$\begin{aligned}&E\{[S^*(T+\Delta)-m(T+\Delta)][m(T+\Delta)-S(T+\Delta)]\}\\&\quad=-E\{[S(T+\Delta)-m(T+\Delta)]\}\\&\qquad\times\sum_{i=1}^{\infty}a_i(S(t)+N(t)-m(t),\varphi_i(t))\\&\quad=-\sum_{i=1}^{\infty}a_i(r_{SN}(T+\Delta,t),\varphi_i(t)),\end{aligned}$$

where, of course,

$$(r_{SN}(T+\Delta,t),\varphi_i(t))=\int_0^T r_{SN}(T+\Delta,t)\varphi_i(t)dt.$$

Finally we obtain

$$\begin{aligned}J&=r_S(T+\Delta,T+\Delta)\\&\quad+\sum_{i=1}^{\infty}\frac{a_i^2}{\lambda_i}-2\sum_{i=1}^{\infty}a_i(r_{SN}(T+\Delta,t),\varphi_i(t)).\end{aligned} \tag{9}$$

Since we assume $m(t)$ to be a polynomial of degree n; we write

$$m(t)=\sum_{j=0}^{n}\beta_j t^j.$$

Since

$$(m,\varphi_i)=\sum_{j=0}^{n}\beta_j(t^j,\varphi_i),$$

the condition given by Eq. (8) becomes

$$\sum_{i=1}^{\infty}a_i\sum_{j=0}^{n}\beta_j(t^j,\varphi_i)=\sum_{j=0}^{n}\beta_j(T+\Delta)^j.$$

Interchanging the order of summation and writing $\alpha_{ij}=(t^j,\varphi_i)$, we obtain

$$\sum_{j=0}^{n}\beta_j\left\{\sum_{i=1}^{\infty}\alpha_{ij}a_i-(T+\Delta)^j\right\}=0.$$

Since this condition must be satisfied for every finite valued $m(t)$ and hence in particular for every linearly independent set of $\{\beta_j\}$, $j=0,1,2,\cdots,n$, we must have

$$\sum_{i=1}^{\infty}\alpha_{ij}a_i-(T+\Delta)^j=0,\quad\text{for } j=0,1,2,\cdots,n. \tag{10}$$

Any unbiased least squares predictor of $S(T+\Delta)$ is obtained by minimizing J(from Eq. (9)) subject to the $n+1$ linear equalities given by Eq. (10.). We will show

[4] U. Grenander, Arkiv Mat. Astron. Fysik., Band 1, Häfte 3, Stockholm (1950).

that there exists a unique set of $\{a_i\}$, $i=1, 2, \cdots$ satisfying these conditions. Introducing Lagrange multipliers $2\mu_0, 2\mu_1, \cdots, 2\mu_n$, our problem is to minimize

$$H=J-2\mu_0\sum_{i=1}^{\infty}\alpha_{i0}a_i-2\mu_1\left\{\sum_{i=1}^{\infty}\alpha_{i1}a_i-(T+\Delta)\right\}-\cdots$$
$$-2\mu_n\left\{\sum_{i=1}^{\infty}\alpha_{in}a_i-(T+\Delta)^n\right\}.$$

The solution is obtained from the equations

$$\frac{\partial H}{\partial a_i}=0 \qquad i=1, 2, \cdots,$$

$$\frac{\partial H}{\partial \mu_j}=0 \qquad j=0, 1, 2, \cdots, n.$$

The first set of equations yields

$$a_i=\lambda_i(r_{SN}(T+\Delta, t), \varphi_i(t))$$
$$+\mu_0\lambda_i\alpha_{i0}+\cdots+\mu_n\lambda_i\alpha_{in}, \quad i=1, 2, \cdots, \quad (11)$$

and the second set yields

$$\sum_{i=1}^{\infty}\lambda_i\alpha_{ij}\{(r_{SN}(T+\Delta, t), \varphi_i(t))$$
$$+\mu_0\alpha_{i0}+\cdots+\mu_n\alpha_{in}\}-(T+\Delta)^j=0. \quad (12)$$

Rewriting this set of equations, we have

$$\mu_0\sum_{i=1}^{\infty}\lambda_i\alpha_{i0}\alpha_{ij}+\cdots+\mu_n\sum_{i=1}^{\infty}\lambda_i\alpha_{in}\alpha_{ij}$$
$$=(T+\Delta)^j-\sum_{i=1}^{\infty}\lambda_i\alpha_{ij}(r_{SN}(T+\Delta, t), \varphi_i(t)) \quad (12a)$$
$$j=0, 1, 2, \cdots, n.$$

Since $\alpha_{ij}\alpha_{ik}<T^{jk}M^2{}_i$, it is clear from Assumption 5 that all of the terms in Eq. (12) are finite. The unique solution for the set $\{\mu_i\}$, $i=0, 1, 2, \cdots, n$ of Eq. (12) substituted in Eq. (11) yields the desired coefficients a_i in the linear predictor

$$S^*(T+\Delta)=\sum_{i=1}^{\infty}a_ix_i.$$

This completes the proof of the theorem. The expected error of the predictor is given by Eq. (9) with the value of a_i substituted from Eq. (11). Clearly the condition that the process be nondeterministic is that $J>0$ for $\Delta>0$. (In the particular case that $S(t)$ and $N(t)$ possess power spectra, a necessary and sufficient condition for this in terms of the covariance functions of $S(t)$ and $N(t)$ has been given by M. Loève (see reference 1), our condition $J>0$ must be related to the necessary and sufficient condition for nondeterminism given in terms of the power spectrum by Paley and Wiener.[5] We have not attempted to establish the relation, however).

It is interesting to note the simple form for the predictor when $ES(t)\equiv 0$. In this case a_i is obtained directly from Eq. (11) (or by differentiating Eq. (9) with respect to a_i) as

$$a_i=\lambda_i(r_{SN}(T+\Delta, t), \varphi_i(t)).$$

Substituting this value in Eq. (9), we obtain

$$J_0=r_S(T+\Delta, T+\Delta)-\sum_{i=1}^{\infty}\lambda_i(r_{SN}(T+\Delta, t), \varphi_i(t))^2. \quad (13)$$

The optimum predictor is given by

$$S^*(T+\Delta)=\sum_{i=1}^{\infty}\lambda_ix_i(r_{SN}(T+\Delta, t), \varphi_i(t)). \quad (14)$$

In the case in which there is no noise present, $N(t)\equiv 0$, Eq. (14) has been obtained by Grenander (see reference 4, p. 269), who attributes the result to Karhunen.

We obtain

$$a_i=\lambda_i\int_0^T r_S(T+\Delta, t)\varphi_i(t)dt=\varphi_i^*(T+\Delta),$$

in which $\{\varphi_i^*(T+\Delta)\}$ have been termed "continued eigenfunctions" for an obvious reason. Since $m(t)\equiv 0$, we see from Eq. (7) that

$$x_i=\frac{Z_i}{(\lambda_i)^{\frac{1}{2}}}.$$

Hence, we obtain

$$S^*(T+\Delta)=\sum_{i=1}^{\infty}\frac{\varphi_i^*(T+\Delta)}{(\lambda_i)^{\frac{1}{2}}}Z_i,$$

the explicit result obtained by Grenander.

It is clear that the condition $J>0$ is violated for some processes. Since we have studied nonstationary processes, we can, for example, obtain a deterministic process by defining $r_S(T+\Delta, T+\Delta)\equiv 0$ for all $\Delta\geq\Delta_0$. Then the process is determined with probability one for all $t\geq T+\Delta_0$. It appears to be difficult to obtain any useful criterion on the covariance function $r_{SN}(s, t)$ which will insure that $J_0>0$ in Eq. (13). A simple sufficient criterion closely allied to Assumption 5 can be obtained in the following manner if we make a slight restriction on the behavior of $r_{SN}(s, t)$. If we assume that $r_{SN}(s, t)$ is for fixed s in $[0, T]$ a function of bounded total variation $V_0{}^T$ in the variable t for $t\epsilon[0, T]$, then we can apply a well-known inequality (see

[5] R. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (American Mathematical Society, New York, 1934).

Hobson[6])

$$\left|\int_0^T r_{SN}(T+\Delta, t)\varphi_i(t)dt - r_{SN}(T+\Delta, 0)\int_0^T \varphi_i(t)dt\right| < M_i V_0{}^T,$$

where

$$M_i = \sup\left|\int_a^b \varphi_i(t)dt\right|$$

for all intervals $[a, b]$ in $[0, T]$.

From this inequality it follows easily that

$$\left|\int_0^T r_{SN}(T+\Delta, t)\varphi_i(t)dt\right|^2 \leq M_i{}^2\{|r_{SN}(T+\Delta, 0)|+V_0{}^T\}^2.$$

Hence applying this inequality to Eq. (13), we see that a sufficient condition that $J_0>0$ is that

$$\sum_{i=1}^{\infty}\lambda_i M_i{}^2 < \frac{r_{SN}(T+\Delta, T+\Delta)}{[|r_{SN}(T+\Delta, 0)|+V_0{}^T]^2} \tag{15}$$

for $\Delta>0$.

(Note that the M_i as defined here is slightly different from the M_i appearing in Assumption 5.)

III. THE OPTIMUM PREDICTOR AS A CONDITIONAL EXPECTED VALUE

The classical theory of linear prediction as developed by N. Wiener,[7] A. Kolmogoroff,[8] and others assumes knowledge only concerning the first and second moments of certain probability distributions. In this case it is natural to use the powerful techniques of least squares as has been done. In some applications it seems plausible to assume that the probability distributions of signal and noise are of definite known types. If this can be done, it is possible to develop a prediction theory based upon the theory of conditional probabilities. On the basis of this theory the optimum predictor is always taken to be the conditional expected value of the quantity to be predicted, given the continuum of values of $S(t)+N(t)$ during a time interval of length T. Although all methods of prediction are based upon arbitrary criteria of what constitutes an optimum predictor, the definition of an optimum predictor in terms of a conditional expected value has many desirable properties when the basic distributions are such that in every case mean coincides with mode. In this case the conditional expected value will for symmetrical distributions be the "center of gravity" of the best region of prognosis.[9]

We wish to point out that in the cases in which signal and noise are Gaussian (this apparently covers the majority of applications) that the predictor which is optimum from the standpoint of the theory of conditional probabilities is identical with the best linear predictor which is obtained from the theory of least squares.

Finally we note that if the signal is a Markoff process, the method of conditional probabilities gives an extremely simple method of obtaining a predictor formula. We will illustrate this point later by an example.

Next we introduce the concepts regarding conditional probability distributions which we will require.[10] Consider n one-dimensional random variables $\xi_1, \xi_2, \cdots, \xi_n$. The probability distribution function of this set of variables is a function of n real variables $x_1, x_2, \cdots, x_n$ given by

$$F(x_1, x_2, \cdots, x_n) = Pr.\{\xi_1 \leqslant x_1, \xi_2 \leqslant x_2, \cdots, \xi_n \leqslant x_n\}.$$

If we limit our consideration here to functions

$$F(x_1, x_2, \cdots, x_n)$$

which are absolutely continuous, the frequency function

$$f(x_1, x_2, \cdots, x_n) = \frac{\partial^n F(x_1, x_2, \cdots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

exists. We define the conditional frequency function of ξ_1 relative to the hypothesis that

$$\xi_2 = x_2, \xi_3 = x_3, \cdots, \xi_n = x_n$$

by

$$f(x_1|x_2, x_3, \cdots, x_n) = \frac{f(x_1, x_2, \cdots, x_n)}{\int_{-\infty}^{\infty} f(\xi_1, x_2, x_3, \cdots, x_n)d\xi_1}.$$

We define also the conditional expected value of ξ_1 as

$$E(\xi_1|x_2, x_3, \cdots, x_n) = \frac{\int_{-\infty}^{\infty} \xi_1 f(\xi_1, x_2, x_3, \cdots, x_n)d\xi_1}{\int_{-\infty}^{\infty} f(\xi_1, x_2, x_3, \cdots, x_n)d\xi_1}.$$

Let us now apply this concept to the simple case of prediction which we considered earlier from the point of view of the theory of least squares. We wish to predict $S(T+\Delta)$, $\Delta>0$, knowing the values of $S(t)+N(t)$ during the time interval $[0, T]$. For the observable coordinates $(x_1, x_2, \cdots)$ we take as before the set defined by Eq. (4).

[6] E. W. Hobson, *Theory of Functions of a Real Variable* (Cambridge University Press, Cambridge, 1927), 3rd Ed., Vol. 1.

[7] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (John Wiley and Sons, Inc., New York, 1949).

[8] A. Kolmogoroff, Bull. Acad. Sci. (USSR) **5**, 3 (1941).

[9] For a development of the theory of regions of best prognosis see Grenander, reference 4, pp. 270–272.

[10] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946), Chap. 22.

The predictor is given by

$$S^*(T+\Delta) = \lim_{n\to\infty} E(S(T+\Delta)\,|\,x_1, x_2, \cdots, x_n)$$

$$= \lim_{n\to\infty} \frac{\int_{-\infty}^{\infty} S(T+\Delta) f(S(T+\Delta), x_1, x_2, \cdots, x_n) dS(T+\Delta)}{\int_{-\infty}^{\infty} f(S(T+\Delta), x_1, x_2, \cdots, x_n) dS(T+\Delta)}, \tag{17}$$

in which $f(S(T+\Delta), x_1, x_2, \cdots, x_n)$ denotes the joint frequency function of the signal at the future time $T+\Delta$ and the n observable coordinates $(x_1, x_2, \cdots, x_n)$ given by Eq. (4).

Let us now obtain the expected squared error of the predictor given by Eq. (17). We have

$$E\{S^*(T+\Delta)-S(T+\Delta)\}^2 = \lim_{n\to\infty} \frac{\int_{-\infty}^{\infty} [S-S^*(T+\Delta)]^2 f(S, x_1, x_2, \cdots, x_n) dS}{\int_{-\infty}^{\infty} f(S, x_1, x_2, \cdots, x_n) dS}. \tag{18}$$

In the event that the signal process is Markoffian, another coordinate system is much more natural and gives an easily computable predictor formula. We recall that the process $S(t)$ is Markoffian if the conditional probability density of $S(T+\Delta)$, $\Delta>0$, given the values of $S(t)$ during the interval $[0, T]$, depends only upon the value $S(T)$. In this case, therefore,

$$S^*(T+\Delta) = \frac{\int_{-\infty}^{\infty} S(T+\Delta) f(S(T+\Delta), S(T)+N(T)) dS(T+\Delta)}{\int_{-\infty}^{\infty} f(S(T+\Delta), S(T)+N(T)) dS(T+\Delta)}. \tag{19}$$

To obtain the optimum predictor for a Markoff process, then, we merely need to know the joint probability distribution of the signal at time $T+\Delta$ and signal plus noise at the time T.

The reader will find it an interesting example to show that if $S(t)$ and $N(t)$ are Gaussian that the predictor given by Eq. (14) by the method of least squares is identical with the predictor formula given by Eq. (17) by the method of conditional probabilities.

IV. THE PREDICTION OF ANY GIVEN LINEAR FUNCTIONAL OF $S(t)$

It is easy (at least in principle) to extend the above results to the general case when we wish to predict

$$L(S) = \int_{-\infty}^{\infty} S(t) dg(t).$$

We require the functions

$$m_L = \int_{-\infty}^{\infty} m(t) dg(t).$$

The covariance function $r_S(T+\Delta, T+\Delta)$ must be replaced by

$$r_L = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_S(t_1, t_2) dg(t_1) dg(t_2).$$

The function $r_{SN}(T+\Delta, t)$ clearly must be replaced by the function

$$r_{LSN}(t) = E\{[L(S)-m_L][S(t)+N(t)-m(t)]\} = \int_{-\infty}^{\infty} r_{SN}(\tau, t) dg(\tau).$$

The basic Eq. (9) then becomes

$$J = r_L + \sum_{i=1}^{\infty} \frac{a_i^2}{\lambda_i} - 2\sum_{i=1}^{\infty} a_i(r_{LSN}(t), \varphi_i(t)), \tag{9a}$$

subject to the restriction

$$\sum_{i=1}^{\infty} a_i(m, \varphi_i) = \int_{-\infty}^{\infty} m(t) dg(t) = \int_{-\infty}^{\infty} \left[\sum_{j=0}^{n} \beta_j t^j\right] dg(t). \tag{8a}$$

Equation (10) becomes

$$\sum_{i=1}^{\infty} \alpha_{ij} a_i - \int_{-\infty}^{\infty} t^j dg(t) = 0, \quad j=0, 1, 2, \cdots, n. \tag{10a}$$

We see, therefore, that the general theory is really only a slight modification of the simple case which we have treated in detail. It is left as an exercise to the reader to generalize the predictor Eq. (17) and the expected error (18) for the general case. In the application of Eqs. (17) and (18) it is fortunate that, if $S(t)$ is Gaussian, then the linear functional $L(S)$ is also Gaussian.

V. COMMENTS ON THE FILTERING PROBLEM

We have obtained an explicit solution to the problem of determining the optimum linear predictor. We ob-

tained it in the form of

$$\sum_{i=1}^{\infty} a_i x_i,$$

where x_i is given from Eq. (4), and the $\{a_i\}$ were obtained as a result of solving the $n+1$ simultaneous linear equations given by Eqs. (12) and substituting the solution in Eq. (11) for the a_i. If the set of eigenfunctions $\{\varphi_i(t)\}$ is a complete orthonormal set, the optimum linear predictor can be written as

$$\int_0^T K(t)[S(t)+N(t)]dt,$$

where

$$K(t)=\sum_{i=1}^{\infty} a_i\varphi_i(t).$$

In most practical applications one is not interested in solving a single prediction problem but in obtaining a particular prediction continuously as time progresses. In other words at the time instant t, we wish to predict some linear functional $S^*(t)$ from knowledge of $S(t)+N(t)$ during the time interval $[t-T, t]$. Since the $\{a_i\}$ in $K(t)$ change with time because of the nonstationarity of signal or noise, we put this fact in evidence by writing

$$S^*(t)=\int_{t-T}^{t} K(t,\tau)[S(\tau)+N(\tau)]d\tau. \tag{20}$$

Only in the event that $S(t)$ and $N(t)$ are quasi-stationary in time will the weight function of the filter in Eq. (20) be of the usual form $K(t-\tau)$. Clearly, in the general case that we are considering in this paper, any system of electrical circuits reproducing Eq. (20) must have time-variable circuit constants. Certain phases of the theory of linear filters with nonstationary random inputs have been investigated by Blanc-Lapierre.[11]

VI. EXAMPLES

The only cases in which we can compare our results with N. Wiener's theory are the cases in which the signal is a Markoff process. The reason for this is obvious when one considers the fundamental integral Eq. (3a). The range of integration $[0, T]$ is the interval during which we observe $S(t)+N(t)$. To consider the case in which signal plus noise have been observed for an infinite time we would have to consider Eq. (3a) with a range of integration $[0, \infty]$. However, as is well known, the theory of eigenfunctions and eigenvalues on which we have based our predictor formula breaks down completely when the range of integration is infinite. In predicting the future value of a Markoff process, however, one obtains the same amount of information about the future by observing the process for a finite time T as by observing for an infinite time. If signal and noise are Gaussian and strictly stationary we should obtain the same predicted value and the same expected squared error as one does by Wiener's theory. We show this by the following simple example: Let $S(t)$ and $N(t)$ be Gaussian, and let $m(t)\equiv 0$. For convenience in notation, we define

$$x_1=S(T+\Delta), \quad \text{and} \quad x_2=S(T)+N(T),$$
$$\sigma_1^2=E[S(T+\Delta)]^2 \quad \text{and} \quad \sigma_2^2=E[S(T)+N(T)]^2,$$

with

$$E\{S(T+\Delta)[S(T)+N(T)]\}=\rho\sigma_1\sigma_2.$$

The joint frequency function of x_1 and x_2 is given by

$$f(x_1, x_2)=\frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \times\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1^2}{\sigma_1^2}-\frac{2\rho}{\sigma_1\sigma_2}x_1x_2+\frac{x_2^2}{\sigma_2^2}\right]\right\}.$$

The predicted value of $S(T+\Delta)$ is given then by Eq. (18) as

$$S^*(T+\Delta)=\frac{\displaystyle\int_{-\infty}^{\infty} x_1\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1^2}{\sigma_1^2}-\frac{2\rho}{\sigma_1\sigma_2}x_1x_2+\frac{x_2^2}{\sigma_2^2}\right]\right\}dx_1}{\displaystyle\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1^2}{\sigma_1^2}-\frac{2\rho}{\sigma_1\sigma_2}x_1x_2+\frac{x_2^2}{\sigma_2^2}\right]\right\}dx_1}$$
$$=\rho\sigma_1x_2/\sigma_2. \tag{21}$$

The predicted value of $S(T+\Delta)$ given by Eq. (19) is simply proportional to the last observed value of signal plus noise; i.e., $S(T)+N(T)$.

An example which N. Wiener gives (see reference 7, p. 65) is a special case of this with $N(t)\equiv 0$ and $S(t)$ stationary. In that case $\sigma_1=\sigma_2$ and $\rho=e^{-\Delta}$ Hence we obtain

$$S^*(T+\Delta)=e^{-\Delta}S(T) \tag{22}$$

in agreement with Wiener's result. The expected error of the predictor is given by

$$J=\frac{\displaystyle\int_{-\infty}^{\infty}\left(x_1-\rho\frac{\sigma_1}{\sigma_2}x_2\right)^2 f(x_1, x_2)dx_1}{\displaystyle\int_{-\infty}^{\infty} f(x_1, x_2)dx_1}$$
$$=\sigma_1^2(1-e^{-2\Delta}).$$

[11] A. Blanc-Lapierre, Compt. rend. **220**, 127 (1946).

Letters to the Editor

The Angular Shift of the Transmitted X-Ray Beam near the Bragg Angle*

L. Carlton Brown, George L. Rogosa, and Guenter Schwarz
Department of Physics, Florida State University, Tallahassee, Florida
(Received June 2, 1952)

RAMACHANDRAN[1] has recently investigated the transmission of x-rays through calcite at settings near the Bragg angle and using the theory of Von Laue[2] he has calculated the transmitted and reflected intensities. Hirsch,[3] Zachariasen,[4] and Ewald[5] have also treated this problem. Experiments have been performed by Borrmann,[6] Campbell,[7] and Schwarz and Rogosa.[8] The latter have shown the thickness dependence of the shape of the transmitted curve at a wavelength of 0.7A for calcite crystals.

Ramachandran and Hirsch have pointed out that for "thick" crystals the theory predicts sharp maxima in the reflected and transmitted beams but that the peaks do not occur at the same angular setting. They also predict an asymmetry of the transmitted curve.

We have measured the intensities of the transmitted and reflected beams and our results support these theoretical conclusions. Mo x-ray radiation was produced by using a highly stabilized high voltage power supply. A double crystal spectrometer served as monochromator. It was set to transmit the peak of the MoKα_1 line. The radiation coming from the second crystal of this monochromator passed over the center of a crystal table which could be rotated in steps of 1 second of arc. A calcite crystal was mounted on the axis of this table. This crystal was adjusted to diffract the monochromatic x-rays from its internal planes giving a minimum width of the rocking curve. At each crystal setting the intensity of the reflected and transmitted beams was measured successively with a Geiger counter. Figures 1 and 2 show the curves for calcite of thicknesses 2.26 mm and 2.56 mm. Defining the Bragg angle θ_B as that angle at which the peak of the reflected curve occurs, one finds that the peak of the transmitted curve is shifted towards smaller angles by about 1 second of arc. The asymmetry of the transmitted curve is also clearly shown with the higher absorption occurring on the side of $\theta > \theta_B$.

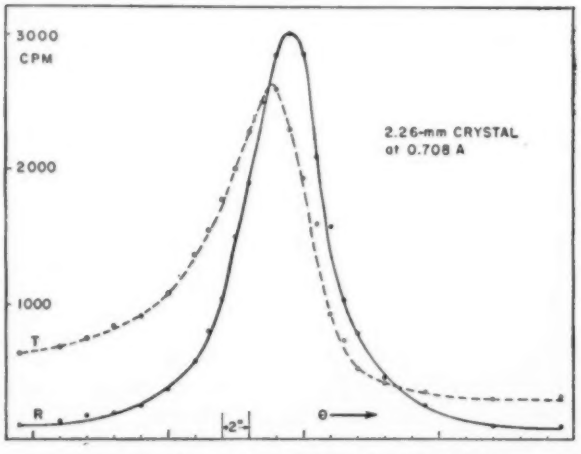


FIG. 1. Transmitted (T) and reflected (R) intensities for a 2.26-mm calcite crystal.

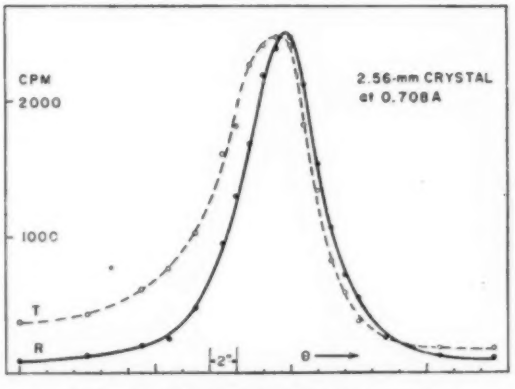


FIG. 2. Transmitted (T) and reflected (R) intensities for a 2.56-mm calcite crystal.

* This work was supported in part by a grant from the Research Corporation.

[1] G. N. Ramachandran, J. Appl. Phys. **23**, 500 (1952); G. N. Ramachandran and Gopinath Kartha, Proc. Indian Acad. Sci. **35**, 145 (1952).
[2] M. Von Laue, Acta Cryst. **2**, 106 (1949).
[3] P. B. Hirsch, Acta Cryst. **5**, 176 (1952).
[4] W. H. Zachariasen, Proc. Natl. Acad. Sci. **38**, 378 (1952).
[5] P. P. Ewald (private communication).
[6] G. Borrmann, Physik. Z. **42**, 157 (1941); Z. Physik **127**, 297 (1950).
[7] H. N. Campbell, J. Appl. Phys. **22**, 1139 (1951).
[8] G. Schwarz and G. L. Rogosa, Phys. Rev. **86**, 421 (1952).

Shadow of an Electron Beam

H. M. Smith and S. Hansen
Electron Tube Laboratories, Hughes Aircraft Company, Culver City, California
(Received June 2, 1952)

THE photograph, Fig. 1. shows the electronic shadow cast by an electron beam. This shadow was first observed during some tests being made on cathode-ray storage devices. The conditions for observing this shadow of the electron beam are shown schematically in Fig. 2.

The fluorescent screen on the front surface of the tube is uniformly illuminated by the flood beam. This flood beam, after being decelerated to approximately 50 volts as it passes through an electrostatic shielding mesh, is then accelerated to approximately 3 kv to produce a reasonably bright fluorescence of the viewing screen. The focused beam which casts the shadow originates from an electron gun located below the axis of the tube. Rough calculations indicate that the space-charge field of the focused beam produces the shadow by deflecting the low velocity flood electrons as they pass near this beam.

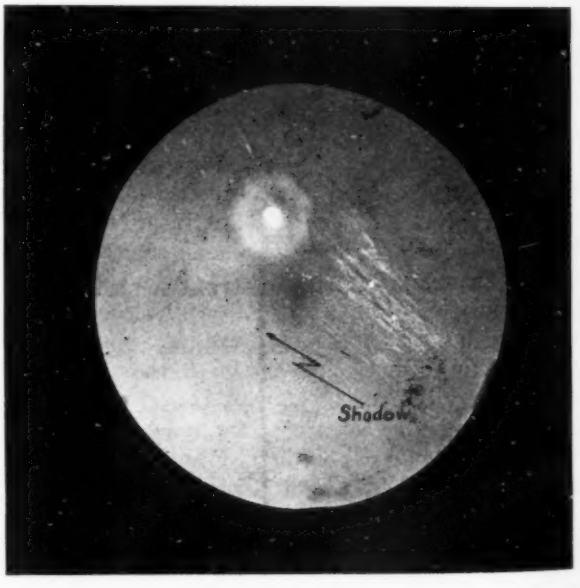


FIG. 1.

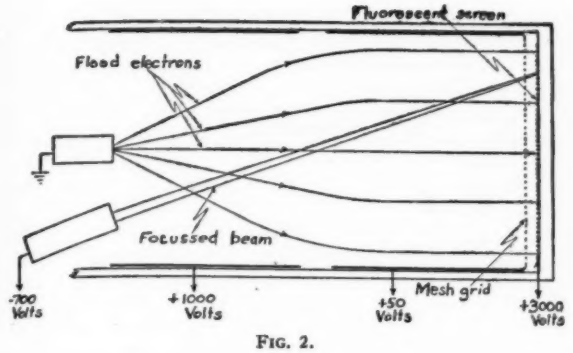


FIG. 2.

In the photograph the focused spot looks enlarged because of overexposure required by the lower background light level in the vicinity of the shadow. Visually, however, the shadow is reasonably sharp, and it is possible to observe the cross-over of the focused beam if it is focused some distance in front of the viewing screen.

Self-Heating Thermionic Vacuum Tubes

E. G. HOPKINS
The New South Wales University of Technology, School of Electrical Engineering, Broadway, Sydney, Australia
(Received June 2, 1952)

EXPERIMENTAL thermionic vacuum tubes in which an alternating electron current passing between two identical planar oxide cathodes provides the heat to maintain the cathodes at their working temperature have been described in a previous letter.[1] Further work has demonstrated the principles of operation of two potentially useful electronic devices in addition to the triode previously described.

The "thermionic rheostat" is a self-heating diode in which the interelectrode spacing is varied to control the power developed in a load connected in series with it to an ac supply. A wide range of control is obtained without exceeding the permissible temperature limits of the oxide cathodes. One experimental tube having a cathode area of 8 cm² varied the power developed in a 240-ohm resistor connected in series with it to a 240-volt supply from 14 watts to 122 watts with a cathode movement of 0.5 cm.

The "thermionic voltage regulator" is a self-heating diode which, when connected in series with a resistor, operates with only a portion of the cathodes heated to the temperature necessary for space-charge limited operation. By considering the conditions necessary for thermal equilibrium it can be proved that, subject to certain assumptions, the voltage across the diode decreases with increasing current until regions of constant temperature are established at the hottest parts of the cathodes. The voltage is independent of current for higher currents. The voltage across an experimental diode having 0.010-cm nickel cathodes 10 cm long and 0.6 cm wide and a spacing of 0.35 cm varied between 98 volts and 100 volts over the range 100 mA to 200 mA. Sudden changes in the supply voltage caused ac transients in the tube voltage having time-constants of 10 sec to 20 sec.

One self-heating diode has run continuously for 6000 hours.

[1] E. G. Hopkins, J. Appl. Phys. **21**, 841 (1950).

The Kirkendall Effect in Alloy Systems

HERBERT N. HERSH
National Advisory Committee for Aeronautics, Lewis Flight Propulsion Laboratory, Cleveland, Ohio
(Received June 12, 1952)

IT was the purpose of the present study to investigate and determine the nature of the forces acting within the diffusion zone which cause the Kirkendall effect. In the usual experiment, inert markers are placed on opposite faces of a rectangular bar of the alloy AB which is then completely covered by a thick layer of B.[1] The Kirkendall shift is given by the displacement of the markers after the diffusion anneal. It has been found that this displacement is proportional to the square root of the time.[1,2] Dimensional changes seem to be restricted to the direction parallel to the diffusion flow. It was desired to demonstrate the Kirkendall effect in a more vivid way which would not involve such problems as the method of joining and the laying down of markers. This was accomplished by heating a homogeneous alloy (69.3 percent copper, 30.7 percent zinc) in a kinetic vacuum space, thereby losing some of the volatile zinc. Since every one of the zinc atoms occupied a lattice site, it is easy to see that their sudden removal may create an immediate excess of vacancies. The existence of these surplus vacancies makes the crystal unstable with respect to them, which gives rise to a virtual force tending to reduce the crystal to its minimum potential energy by equilibrating their number. The mechanism for this is not known, but may be described as some sort of atomic "sintering." Physically, the copper atoms are in the lattice at nonequilibrium distances for their new concentration. The additional vacancies are destroyed by migrating to an external surface or by coalescing to form macroscopic pores.

TABLE I. Dimensional changes upon dezincification.

Specimen	Distance to opposite faces (mm) for given heating times (hr) 0	24	47	70	134	Shrinkage (mm) for given heating times (hr) 24	47	70	134
DK-1a	16.228	16.201	16.199	16.192	16.180	0.027	0.029	0.037	0.049
DK-2a	13.825	13.811	13.800	13.795	13.782	0.014	0.025	0.030	0.043
DK-3a	12.940	12.922	12.916	12.908	12.898	0.018	0.024	0.032	0.042
DK-1b	7.407	7.393	7.381	7.376	7.368	0.014	0.026	0.031	0.039
DK-2b	7.285	7.270	7.259	7.264	7.249	0.015	0.024	0.031	0.036
DK-3b	7.392	7.371	7.362	7.361	7.350	0.021	0.030	0.031	0.042
DK-1c	8.118	8.101	8.097	8.093	8.088	0.017	0.021	0.025	0.030
DK-2c	7.680	7.661	7.655	7.650	7.645	0.019	0.025	0.030	0.035
DK-3c	8.325	8.310	8.301	8.300	8.292	0.015	0.024	0.025	0.033

The specimens were dezincified in a quartz high vacuum system by means of a Wheelco-controlled tube furnace at about 700°C. The dimensions of the specimens were determined after various periods of dezincification and are given in Table I together with the dimensional changes. The data of Table I show that brass

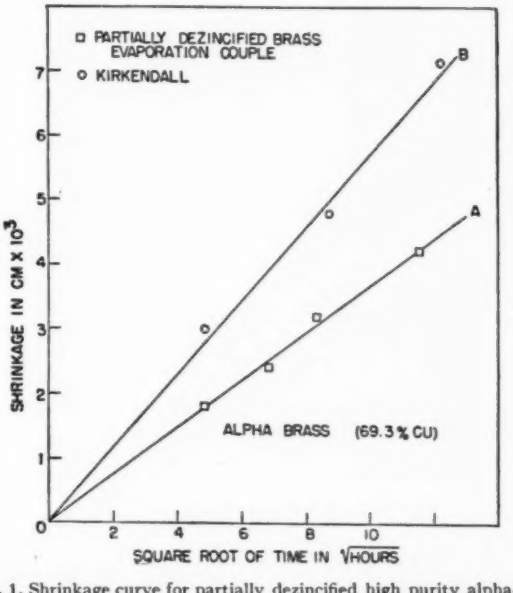


FIG. 1. Shrinkage curve for partially dezincified high purity alpha-brass heated in vacuum as a function of the square root of the heating time (curve A) showing conformity to the diffusion law $x^2 \sim t$. Curve B shows the shrinkage curve in a standard copper-brass-copper diffusion couple (after Kirkendall) (see reference 1) which follows the same law.

contracts upon dezincification. At the temperature of the run, no dimensional changes occur in a similar specimen of pure copper. Thus, the shrinkage is not due merely to a loss of mass of the sample as a whole, but is due to the ability of the copper to densify in some fashion. A plot of the data of Table I shows that the movement of the solid interface varies as the square root of the time, as did Kirkendall's markers (Fig. 1). Photomicrographs of the interior of some of the specimens were made. These showed a large amount of porosity with none at the center. Some of the pores were lined up along grain boundaries. The porosity, which is generated to various extents in other systems upon rapid removal of the volatile component, in the case of the present system indicates that maximum shrinkage did not occur. The copper-zinc system was the only one in which porosity was noticed after diffusion by Correa da Silva and Mehl, but the recent experiments of Baluffi and Alexander,[3] as well as results in this laboratory, have shown that porosity is usually developed in interdiffusion in other systems. It is noted that the contractions are not limited to one direction. Contractions, such as occur during ordinary diffusion, appear to be unidirectional as a result of the experimental configuration, but are essentially volume effects in the region of the diffusion zone. This is shown quite clearly with narrow lengths of sheet brass; one specimen, originally 5.0 cm long, became 4.6 cm in length after an evaporation treatment of one and one-half days at 700°C. In these thin specimens porosity pervades the whole volume. With the larger specimens, used in compiling the data of Table I, the dimensional changes are roughly independent of the sample size within the experimental error. This is because the contracting mechanism operates within the surface zone and not throughout the whole solid.

[1] A. D. Smigelskas and E. O. Kirkendall, Trans. Am. Inst. Min. Met. Engrs. 171, 130 (1947).
[2] L. C. Correa da Silva and R. F. Mehl, J. Metals 188, 1219 (1950).
[3] R. W. Baluffi and B. H. Alexander, J. Metals 4, 152 (1952).

An Attempt to Explain a Discrepancy in Observations of the Heat Transfer between a Vertical Surface and a Fluidized Bed of Particles

MAX JAKOB
Illinois Institute of Technology, Chicago, Illinois
(Received June 13, 1952)

RECENTLY Dow and Jakob published[1] measurements of the heat transfer on the vertical surface of cylindric fluidized beds composed of air with different solids, and expressed their results by an equation, according to which

$$(N_{Nu})_t \sim (N_{Re})_t^{0.80} (D_t/D_p)^{0.17}, \quad (1)$$

where $(N_{Nu})_t$ is a Nusselt number based on the tube diameter D_t as characteristic length, $(N_{Re})_t$ is a Reynolds number, also based on D_t and on the velocity which the gas would have had if no solid particles were present, and D_p is the particle size, for instance, an equivalent particle diameter.

Van Heerden, Nobel, and van Krevelen,[2] on the other hand, in experiments with different gases and solids, found

$$(N_{Nu})_p \sim (N_{Re})_p^{0.45}, \quad (2)$$

where the subscript p refers to D_p as characteristic length.

The relation (1) can readily be transformed to

$$(N_{Nu})_p \sim (N_{Re})_p^{0.80} (D_t/D_p)^{-0.63}, \quad (3)$$

where

$$(N_{Nu})_p = (N_{Nu})_t (D_p/D_t)$$

and

$$(N_{Re})_p = (N_{Re})_t (D_p/D_t).$$

For a fixed value[3] of D_t/D_p there remains the peculiar discrepancy between the exponents of $(N_{Re})_p$ in the relations (2) and (3). After searching in vain for any other possible reasons[4] the writer came to the conclusion that the direction of the heat transfer may cause this enormous difference. Dow and Jakob heated the fluidized bed, van Heerden and associates cooled it from outside. Usually there is no appreciable difference in the exponent of N_{Re} in heating and cooling experiments. In fluidization, however, there may occur an essential difference in the mechanism of heat transfer on the surface.

Because of the small velocity of the bulk of the gas a superimposed free-convection effect may be relatively considerable. Both our and the Dutch workers' corresponding experiments would have been in the range of laminar flow if no particles had been present. The free convection would have made the velocity gradient steeper close to the surface in our case, less steep in their case. Therefore, the particles are to meet upward velocities of the gas, small enough to allow their falling down, at a larger distance from a cooling wall than from a heating wall. Hence, the interference with the boundary layer, to which we attributed the large heat transfer[1] observed by us, is likely to occur on a heating surface while the particles may not be driven so closely to a cooling surface as to considerably disturb the boundary layer.

So, it may be understood that the exponent found by van Heerden and associates, was between the exponent 1/3 which would have to be expected in laminar flow[5] and 0.8, often observed in turbulent flow.

In other words, turbulence conditions may be obtained in heating, but not attained in cooling, because of the difference in the interaction of forced and free convection of the fluid with the fall of particles along the surface of heat transfer. Considerably beyond the range of our experimentation, $(N_{Re})_p \approx 0.2$ to 2.9, other features of the mechanism of convection seem to prevail.

[1] W. M. Dow and M. Jakob, Chem. Engg. Progr. 47, 637 (1951), paper presented at the Heat Transfer and Fluid Mechanics Institute, Los Angeles, California, June, 1950.
[2] van Heerden, Nobel, and van Krevelen, Chem. Engg. Science 1, 51 (1951).
[3] This ratio was varied between 305 and 1104 in Dow and Jakob's experiments and between 131 and 1083 in the work of van Heerden *et al.*
[4] Dr. van Heerden in a letter to me, considered the possibility that our work was performed in a range in which his experimental data would also indicate a higher exponent. However, I found that, after proper transformation, our range was just the one in which van Heerden and associates obtained the lower exponent.
[5] See, for instance, Eq. (26-5) in M. Jakob, *Heat Transfer* (John Wiley & Sons, Inc., New York, 1949), Vol. I.

Some Variational Principles for Problems in Transmission Lines

MIKIO NAMIKI AND HIROSHI TAKAHASHI
Department of Applied Physics, Waseda University, Tokyo, Japan
(Received June 16, 1952)

IN treatment of the transmission line we should solve the following equations under the appropriate boundary condition:

$$(dV(x)/dx) + Z(x)I(x) = 0, \quad (dI(x)/dx) + Y(x)V(x) = 0, \quad (1)$$

where $V(x)$ and $I(x)$ are the voltage and current and $Z(x)$ and $Y(x)$ are the impedance and admittance per unit length of the line at a point x. For the most usual cases it is sufficient to obtain the input impedance ζ for the given load impedance R. This requires $(V/I)_{output} = R$ as the boundary condition. However, we cannot solve analytically this problem for general functions $Z(x)$ and $Y(x)$, although there are simple solutions for constant Z and Y. In this letter we shall formulate some variational principles for the above problem.

Now we consider the next functional of the function V and I:

$$J_{VI} = \frac{V(a)}{I(a)} + \frac{1}{I^2(a)} \int_a^b \left\{ I\left(\frac{dV}{dx} + ZI\right) - V\left(\frac{dI}{dx} + YV\right) \right\} dx, \quad (2)$$

where a and b are input and output point, respectively. If V and I are solutions of Eqs. (1) with $(V/I)_{x=b} = R$, J_{VI} is equal to the desired input impedance ζ. Taking the first variation of J_{VI} due

to the independent variation of V and I,

$$\delta J_{VI}=\frac{1}{I^2(a)}\int_a^b\left\{\delta I\left(\frac{dV}{dx}+ZI\right)-\delta V\left(\frac{dI}{dx}+YV\right)\right\}dx$$
$$-\frac{2\delta I(a)}{I^2(a)}\int_a^b\left\{I\left(\frac{dV}{dx}+ZI\right)-V\left(\frac{dI}{dx}+YV\right)\right\}dx+I^2(b)\delta\left(\frac{V}{I}\right)_{x=b}.$$

Since the last term vanishes owing to the boundary condition, the Eulers equations of J_{VI} are Eqs. (1). Thus the functional J_{VI} is a variational expression for the input impedance ζ under the above boundary condition.

Secondly, we consider variational expressions written by one independent function I or V. Eliminating V by the second equation of Eqs. (1), we get a second-order differential equation of I instead ϕ of Eqs. (1), that is,

$$L(I)=0, \quad L\equiv(d/dx)(1/Y)(d/dx)-Z. \tag{3}$$

Then a variational expression for ζ is also given by the functional

$$J_I=-[(dI/dx)/YI]_{x=a}-(1/I^2(a))\int_a^b IZ(I)dx, \tag{4}$$

on which the boundary condition $-[(dI/dx)/YI]_{x=b}=R$ should be imposed. After integration by parts we can rewrite Eqs. (4), namely,

$$J_I'=\frac{1}{I^2(a)}\left[\int_a^b\left\{\frac{1}{Y}\left(\frac{dI}{dx}\right)^2+ZI^2\right\}dx+RI^2(b)\right]. \tag{5}$$

It is to be noticed that J_I and J_I' have somewhat different effects for approximation although the both expressions are mathematically identical. Variational expressions written by V have the equation obtained by eliminating I from Eqs. (1) as its Euler's equation, and one of them is considered as follows:

$$J_V=\frac{V^2(a)}{\int_a^b\{(1/Z)(dV/dx)^2+YV^2\}dx+(1/R)V^2(b)}, \tag{6}$$

on which the boundary condition $-[ZV/(dV/dx)]_{x=b}=R$ is imposed. There is an alternative variational expression derived from J_V,

$$J_V'=\zeta'[2-(\zeta'/J_V)], \tag{7}$$

where

$$\zeta'=-[ZV/(dV/dx)]_{x=a}.$$

The above-mentioned expressions are useful to approximate computation of ζ because of their stationary charactor. For instance, if we use the arbitrary functions with deviations of the order of magnitude ϵ from the exact solutions as trial function, we can get the impedance ζ with the errors of the order of magnitude ϵ^2. In order to see the charactor of the above expressions more concisely, we consider first the case in which R, Z, and Y are real and positive numbers. In this case J_I' and J_V have definite charactor for the approximation of ζ. Let I_0 and V_0 be the exact solutions and let $\Delta I=I-I_0$ and $\Delta V=V-V_0$ be entirely arbitrary functions. Because of the positive definite charactor of $J_I'[\Delta I]$ and $J_V[\Delta V]$, it is easily shown that

$$\Delta\zeta_I=J_I'[I]-J_I'[I_0]\geqslant 0,$$
$$\Delta\zeta_V=J_V[V]-J_V[V_0]\leqslant 0.$$

Hence

$$J_I'\geqslant\zeta\geqslant J_V.$$

Thus we can obtain the upper and lower limits of ζ by using J_I' and J_V simultaneously. These characters are almost conserved with suitable trial functions for more general cases, although there are, of course, many exceptional cases. Consequently, we can perform the computation of the input impedance with good approximation even for the considerably general distributions of Z and Y. For example, we obtained very good approximations by the simplest functions $1+Ax$ or $\exp(\alpha x)$ for the constant line or exponential line. The relation of J_V' to the exact ζ is similar to that of J_V in the vicinity of the stationary point, but the stationary character of J_V is better than that of the J_V'. There is, however, a possibility of better approximations by using J_V', if trial functions are specially chosen. The expression J_{VI} written by two independent functions V and I has entirely different relation to the exact ζ. Although there is also a possibility of better approximations by J_{VI}, it is necessary to take some care for calculations. Thus, it is most desirable to use the above mentioned variational expressions with suitable choices and combinations.

Finally we consider a variational method for the determination of a circuit parameter from given ζ and R. Suppose that we can put

$$Z=Z_0+Bv, \quad Y=Y_0+Bw,$$

where Z_0, Y_0, v, and w are given functions and B is an unknown parameter which is to be determined by given ζ and R. Under the boundary conditions $(V/I)_{x=a}=\zeta$ and $(V/I)_{x=b}=R$, the functional

$$J_B=-\frac{\int_a^b\{I[(dV/dx)+Z_0I]-V[dI/dx)+Y_0V]\}dx}{\int_a^b\{vI^2-wV^2\}dx} \tag{7}$$

is a variational expression for the determination of the parameter B.

The above mentioned methods with some extensions will also be useful for the phenomenological treatment of the distributed coupled circuits. Moreover, we can construct similar expressions for the four constants A, B, C, and D of the transmission line looked upon as a four-terminal network. The details will be published hereafter.

Metallic Corrosion Influenced by Ultrasonic Waves

SHIGETO YAMAGUCHI
Scientific Research Institute, Ltd., Bunkyo-ku, Tokyo, Japan
(Received June 16, 1952)

IT is known that the corrosion of iron, zinc, aluminum, magnesium, etc. is slower in moving water than in motionless water. This phenomenon was explained by the protective oxide film theory,[1] i.e., the supply of oxygen to metal faces is more adequate in moving water for the formation of complete films than in water at rest. In the present study, ultrasonic waves were employed to stir the corroding solution in which immersion tests were carried out. From the results obtained here, the wet corrosion of metals was more reasonably elucidated by colloid science than by the oxide film theory.

The test pieces here used were two disks (diameter: 7 mm, thickness: 3 mm) of magnesium polished coarsely with emery papers. One was subjected to corrosion in a saline water at rest (NaCl: 3 percent) for 5 hours at 20 degrees C. The other was immersed in a saline water of the same concentration as in the case of the first test piece, but this time ultrasonic waves (frequency: about 1 megacycle per second; acoustical intensity from transducer: about 2.5 watts/cm²; the direction of waves was approximately perpendicular to the specimen face) were applied to the immersed test piece for 5 hours at 20 degrees C. It was noted here that the generation of hydrogen gas was more violent in the liquid at rest than in the liquid agitated by ultrasonic waves. The corroded surfaces of the two specimens are shown in Figs. 1 (A) and (B). Figure 1(A) of the specimen corroded in the unagitated liquid shows a pitted surface covered with a thick precipitate of corrosion products. On the contrary, in Fig. 1 (B) traces of polish marks on the substrate are found as well as corrosion products. Since the test piece of Fig. 1 (B) shows its surface to be barely corroded, it is evident that the corrosion under the influence of ultrasonic waves is slower than that in the undisturbed liquid. When the energy of ultrasonic waves was increased to about 5 watts/cm², the generation of hydrogen gas became rather marked. This was interpreted to be due to the fact that the corrosion products were being removed from the substrate under the too violent shock (corrosion-erosion). The phenomenon of cavitation was not distinctively recognized in a short time (about 1 minute) of agitation with the acoustical intensity of 5 watts/cm². It is,

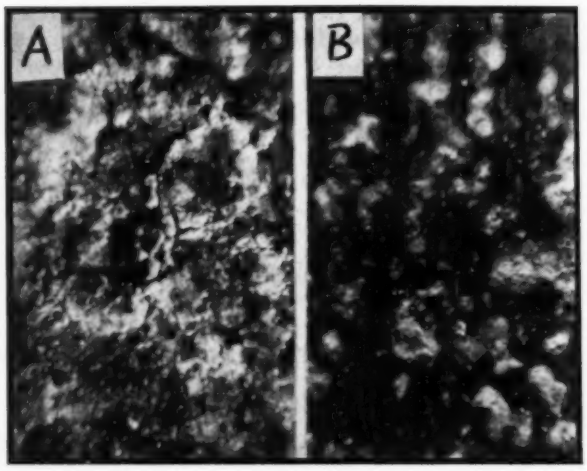


FIG. 1. (A) The surface corroded in unagitated saline water. (B) The surface corroded under the shock of ultrasonic waves. The corrosion of the surface (A) is much more severe than that of the surface (B). ×20.

however, considered that cavitation takes place in a longer time of agitation and with a larger intensity of ultrasonic waves.

It is already known by electron diffraction that the corrosion products of magnesium in a saline water are magnesium oxide and hydroxide.[2] These substances form a colloidal suspension in water which show thixotropic phenomenon. It is known that the coagulation rate in thixotropy is retarded by the mechanical shock of ultrasonic waves.[3] The corrosion of magnesium observed in the present study might be elucidated as a thixotropic phenomenon taking place on the metal faces. The colloidal particles of corrosion products will remain so fine that they can form a compact layer able to protect the substrate from corrosion, so far as a moderate thixotropic shock is given to the corrosion products by ultrasonic waves. Since the corrosion products are easily coagulated in an unagitated saline water, they will form a coarse and porous layer unable to protect the metals from further corrosion. It might be explained as a thixotropic phenomenon why the corrosion of iron, zinc, and aluminum takes place more slowly in moving water than in still water.

The fact that the metallic corrosion is much influenced by ultrasonic waves suggests that the colloidal products of corrosion exhibit thixotropic phenomenon on metal faces.

In the standard corrosion tests for materials, there is one where the test is carried out in mechanically agitated corroding media.[4] This agitation can advantageously be replaced by the shock of ultrasonic waves, because the latter is more quantitative than the former.

[1] U. R. Evans, Nature **168**, 853 (1951).
[2] S. Yamaguchi, Nature **146**, 654 (1940); Japan Science Rev. **1**, 24 (1949).
[3] B. Jirgensons and N. Straumanis, *Kurzes Lehrbuch der Kolloidchemie* (Springer-Verlag, Berlin, 1949), p. 250.
[4] W. Wiederholt, Werkstoff u. Korrosion **9**, 334 (1951).

Significance of Flow-Patterns for Initial Convection in Porous Media

H. L. MORRISON AND F. T. ROGERS, JR.
*U. S. Naval Ordnance Test Station, Inyokern, California**
(Received June 23, 1952)

A RECENT series of papers[1] has discussed several aspects of the initiation of convective flow of a liquid (heated from below) in a porous medium. Major emphasis has been given to the process as it occurs in real liquids rather than in ideal ones and under slowly changing thermal environments rather than static ones; in particular, attempts have been made to explain the great ease with which the process can be made to occur, relative to predictions of extant theory. We wish here briefly to mention some later findings as to both experiment and theory, in clarification of the state of present knowledge.

With the advice of Dr. R. H. Olds we have obtained some numerical solutions to extant theory[2] by means of the N.O.T.S. analog computer.[3] Also, following up a suggestion by Professor L. M. Grossman, we have recast this theory as a variational problem.[4] Each of these methods represents a great advance in rigor over those which have been published, and both yield solutions which are in excellent general agreement with each other. Thus it is clear that (in the notation of reference 2) the present formulation of theory properly leads to $\nu\beta_c D$ generally higher for given νBD than does the classical theory,[5] to $\nu\beta_c D$ as a double-valued function of νBD, and to $\partial(\nu\beta_c D)\partial(\nu BD)$ infinite at a finite value of νBD. These results are in sharp contradistinction to experimental experience, and emphasize that the *calculations in reference 2 are only approximate even though they correlate extremely well with all available data.*

Inherent in past formulations of the theory is the assumption of a definite class of initial-flow patterns, namely, one in which $(\Omega_{xz}+\Omega_{yy})/\Omega$ and $(\theta_{xz}+\theta_{yy})/\theta$ are taken equal to the negative of a positive constant, $A=l^2+m^2$. Such flow patterns would be of the cellular types which are known experimentally for convection not in porous media. Though such flows may occur in porous media, no experimental observations of them have been reported, and in fact the above remarks suggest that flow patterns for all published data[1] may have been of a quite different character. Quite recently one of us (HLM) has been able directly to observe flow patterns, using a new technique the details of which will be published later. Although not all features of these patterns have been identified quantitatively, it is clear that they are not of the conventional cellular types but appear to be physically much more simple. These observations suggest a greater similarity than had been suspected, as between liquid convection in porous media and those processes reported by O. G. Sutton for convection of gases not in porous media. In his paper,[6] Professor Sutton describes a "columnar" flow, and our observations seem to be in better accord with this type than with the cellular.

The classical theory[5] may be modified to take some account of these findings in the following way. The fundamental laws are $\mu\mathbf{v}=-k(\nabla\Omega-\mathbf{g}\rho_0\alpha\theta)$, $\nabla\cdot\mathbf{v}=0$, and $h^2\nabla^2\Theta=\dot{\Theta}+\mathbf{v}\cdot\nabla\Theta$. Using cylindrical coordinates (z, r, φ) with $\mathbf{z}$ directed upward along a flow "column," and considering all processes independent of φ, the first two can be made to yield $w_z-k(r\Omega_r)_r/r\mu=0$ and the last to yield $h^2[\theta_{zz}+(r\theta_r)_r/r]=\dot{\theta}+wT_z$. The modification consists of using $(r\Omega_r)_r/r=-E^2\Omega$ and $(r\theta_r)_r/r=-E^2\theta$ rather than the relations given in the above paragraph; this amounts to the assumption that both Ω and θ are small and proportional to $Z_0(Er)$, where Z_0 denotes the Bessel's functions of zero order, rather than proportional to $\sin lx \sin my$. Thus, by reference again to the first law, there results the differential system

$$\left.\begin{aligned}(\mu w_z)_z/\mu-E^2w+E^2kg\rho_0\alpha\theta/\mu=0\\ \theta_{zz}-E^2\theta-T'w/h^2=0\end{aligned}\right\}. \tag{1}$$

For constant viscosity and for $T=-\beta_c z$, which restrictions were applied in reference 5, and for the boundary conditions imposed in reference 5, these yield

$$\beta_c=(h^4\mu/kg\rho_0\alpha)[(s^2+E^2)^3/E^2] \tag{2}$$

for the (negative) mean thermal gradient at the onset of convective flow. Now s may be supposed to take on the least of its possible values, $0, \pi/D, 2\pi/D, \cdots$, and E the value $2\pi/Wa_0$, where W is a diameter of "columnar cells" and a_0 is the first zero of J_0; then

$$\beta_c=(4\pi^2h^2\mu/kg\rho_0\alpha D^2)\cdot(D^2/W^2a_0^2). \tag{3}$$

As applied to almost all of the published data,[1] in which $W\sim D$, Eq. (3) predicts critical gradients *many times smaller* than does the classical result, $4\pi^2h^2\mu/kg\rho_0\alpha D^2$. To this extent, the new

observations of flow-patterns, through the above type of modification of theory, appear to point the way to a much more nearly valid understanding of available data.

* Post Office: China Lake, California.
[1] Horton, Morrison, Rogers, and Schilberg, J. Appl. Phys. **16**, 367 (1945); **18**, 849 (1947); **20**, 1027 (1949); **21**, 1177 (1950); **22**, 233 (1951); Nucleonics **9**, 47 (1951); and J. Appl. Phys. **22**, 1476 (1951).
[2] Rogers, Schilberg, and Morrison, J. Appl. Phys. **22**, 1476 (1951).
[3] Rogers and Morrison, Bull. Am. Phys. Soc. **27**, No. 4, 10 (1952).
[4] F. T. Rogers, Jr., Phys. Rev. **85**, 770 (1952).
[5] C. W. Horton and F. T. Rogers, Jr., J. Appl. Phys. **16**, 367 (1945); E. R. Lapwood, Proc. Cambridge Phil. Soc. **44**, 508 (1948).
[6] O. G. Sutton, Proc. Roy. Soc. (London) **A204**, 297 (1950).

The Separation of Stacking Fault Broadening in Cold-Worked Metals*

B. E. Warren and B. L. Averbach
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 8, 1952)

In the cold-working of face-centered cubic metals, slip on the (111) planes may leave mistakes in the regular *ABCABC* sequence. These mistakes are called stacking faults, and Barrett[1] has suggested that they may be responsible for part of the broadening of an x-ray reflection. To separate the effect of stacking faults from the other sources of broadening, Bragg[1] has suggested that there should be some reflections which are not influenced.

If a_1, a_2, a_3 are the cubic axes, the vectors from an atom to two neighboring atoms in the (111) plane are $(a_1-a_3)/2$ and $(a_2-a_3)/2$. One-third the sum of these is the vector displacement to interchange A, B, C positions. $\delta=(a_1+a_2-2a_3)/6$. Owing to this displacement, the parts of the crystal on either side of the slip plane may scatter out of phase, the phase factor being

$$\exp[2\pi i\delta\cdot(hb_1+kb_2+lb_3)]=\exp\{2\pi i[(h+k-2l)/6]\}.$$

The slip plane may be anyone of the eight possibilities $(\pm1, \pm1, \pm1)$, and on each plane there are three possible directions of slip, so that for the two parts on either side of a stacking fault plane the general expression for the phase factor is

$$\exp\{2\pi i[(\pm h\pm k\pm 2l)/6]\},$$

where the 2 can be associated with either the h, the k, or the l.

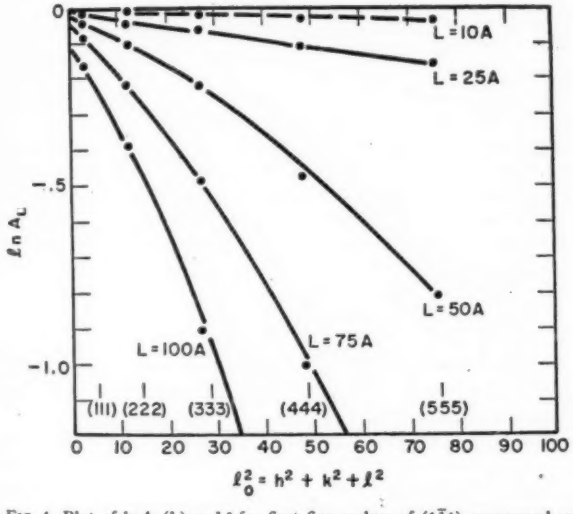


Fig. 1. Plot of $\ln A_n(l_0)$ *vs* l_0^2 for first five orders of $(1\bar{1}1)$, measured on a single crystal of Cu–2 percent Si which has been rolled to 50 percent reduction.

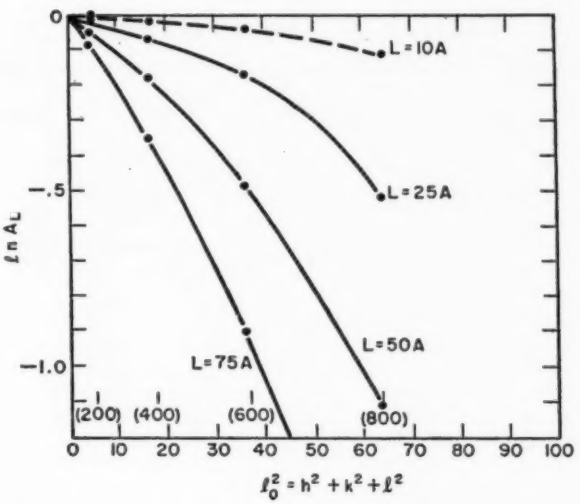


Fig. 2. Plot of $\ln A_n(l_0)$ *vs* l_0^2 for first four orders of (002) measured on a single crystal of Cu–2 percent Si which has been rolled to 50 percent reduction.

It is evident that the phase difference vanishes for the two reflections (333) and (600), and hence stacking faults will produce no broadening in these reflections. The existence of stacking fault broadening can be determined by comparing these two reflections with others where broadening will be produced. On a powder pattern it is not possible to observe (333) and (600) since the third order of a cubic powder reflection always coincides with another reflection, in this case (333) with (511) and (600) with (442). However, by using a single crystal or a sample with very high preferred orientation it is possible to measure the (333) and (600) reflections.

It has been shown[2] that the broadening by either small particle size or distortion can be separated by measuring several orders of a reflection $(00l)$, expressing the peak shape in terms of Fourier coefficients $A_n(l)$, and plotting $\ln A_n(l)$ *vs* l^2. Figures 1 and 2 show such plots for a single crystal of Cu-2 percent Si which has been rolled on the 110 plane in the $\bar{1}12$ direction to 50 percent reduction. Figure 1 shows the Fourier coefficients for 5 orders of $(1\bar{1}1)$, and Fig. 2 shows 4 orders of (002). The quantity $\ln A_n(l_0)$ is plotted against $l_0^2=h^2+k^2+l^2$, and curves are given for a few coefficients, each curve being labeled by a distance $L=na_3$. Since the (333) and (600) points fall on smooth curves through the other reflections, it is evident that in this sample there is no measurable broadening by stacking faults.

The three commonly discussed sources of broadening in cold-worked metals are conveniently separated by plots such as Figs. 1 and 2; since particle size broadening is represented by the intercept on the axis or ordinates, distortion broadening by the slopes of the curves, and stacking fault broadening by the departure of the (333) and (600) points from the smooth curves. Although these results prove that there is no appreciable stacking fault broadening at 2 percent Si, the same sort of measurements should be made on a sample of composition around 5 percent Si where Barrett's results suggest that stacking faults may be important.

* Research sponsored by the AEC under Contracts AT(30–1)–858 and AT(30–1)1002.
[1] C. S. Barrett, *Imperfections in Nearly Perfect Crystals* (John Wiley and Sons, Inc., New York, 1952), Chap. III.
[2] B. E. Warren and B. L. Averbach, J. Appl. Phys. **23**, 497 (1952).

Books Reviewed

The Oxide-Coated Cathode; Volume I Manufacture, Volume II Physics

By G. Herrmann and S. Wagener. Volume I. Pp. 148+viii, Figs. 78. Volume II. Pp. 311+xiv. Figs. 154. Chapman & Hall, Ltd., London.

Die Oxydkathode first appeared as a German publication bearing the date 1943. Most of the first editions found in this country were reprinted by Edwards Brothers in 1946. Although this original edition presented a comprehensive compilation of oxide cathode researches through 1941, the books were unfortunately out-of-date before they became generally available. The results of wartime research and the stimulus that these gave to postwar activity and thinking in this field were conspicuously lacking.

The new English translation of these volumes by Wagener has corrected this deficiency and is well referenced with publications through 1950. The revised volumes total 442 pages as compared to 362 in the original set and contain no less than 232 figures. Since the field of oxide cathodes is of rather restricted interest one might first question the necessity of a two-volume set on this subject. Whereas volume I is concerned exclusively with oxide cathode technology, volume II is devoted to a basic understanding of the oxide cathode. One-half of the latter volume deals with the principles of solid state and physical electrons to establish a basis on which oxide cathode experiments may be discussed. This section on theory (Vol. II, pp. 1–149) is of general interest as a comprehensive review of the energy band structure of solids, electron emission from metals and phenomena in ionic solids. The authors were wise to include this review dealing with the principles to be used later in their treatment of the oxide cathode.

Volume I begins with a short history of oxide cathode development. Cathode manufacture and processing techniques, which for many years were regarded as an art and subject to trade secrets, are more fully discussed than in any single previous publication. The methods of measurement of temperature and emission and characteristic values thereof are presented. This volume closes with a short section on special cathode types. For the cathode technologist Volume I presents an excellent summary of useful information; however, the material is sufficiently specialized that a reader having only a general interest in the subject may begin with Volume II without serious loss.

The first chapter of Volume II treats the thermionic emission of electrons from metals in a rather comprehensive manner. Beginning with a discussion of energy levels and energy bands, the Fermi-Dirac distribution function is derived and applied to a derivation of the Richardson equation for electron emission. The effects of retarding fields, space charge, and accelerating fields on this electron emission are presented. A discussion of atomic film emitters and the temperature dependence of work function complete this chapter. Sixteen ways of measuring the work function of a metal are described in Chapter 2. The authors clearly point out the often neglected fact that five differently defined work functions can result from these sixteen methods. Subjects treated under "phenomena in ionic solids" include: formation of lattice defects, diffusion, ionic conduction, and electronic conduction. Equations are derived for the temperature dependence of the Fermi level and the dependence thereon of electrical conductivity. The latter half of this volume describes and attempts to interpret a large mass of oxide cathode phenomena. These include physical structure, cathode activation, thermionic emission constants, and conductivity measurements. Most of these phenomena are amenable to the type of interpretation that is presented; however, it must be remembered that the field of oxide cathodes is still very active and many controversial issues have not yet been settled.

The text contains an appreciable number of errors which are annoying, as the following list from Chapter 2, Volume II will indicate. Page 73, the slope of a Richardson line is $-5040\psi/T$; it should be -5040ψ. Page 74, the Richardson line gives a value of the work function at $T=0$, this statement results from an assumed temperature dependence of work function whose validity at $T=0$ is to be questioned. Page 85, reference is made to cathode C of Fig. 53; this figure taken from the German text shows no C but instead a "kathode" K. Page 93, Eq. 108 requires the differential dE on the right side.

It is the opinion of the writer that these volumes will find rather wide use as a review of both oxide cathode principles and technology and as a concise summary of solid state electronics.

Albert S. Eisenstein
University of Missouri

Studies in Large Plastic Flow and Fracture

By P. W. Bridgman. Pp. 362+x, Figs. 169. McGraw-Hill Book Company, Inc., New York, 1952. Price $8.00.

In this book Professor Bridgman presents a coordinated description of his work in the field of large plastic deformation and fracture of solids, most of which was done during the past twelve years. A large part of the work described here was previously available only in the form of photostat copies of government documents, and a small amount is material not previously published. The research reported in the book is exclusively that of the author.

Some of the topics covered are: simple longitudinal tension of specimens with a detailed analysis of the stress distribution at the neck; two-dimensional tension; the collapse of hollow cylinders under external pressure; two-dimensional compression; torsion combined with simple compression; and the effects of prestraining the specimen. By subjecting the specimen to a uniform hydrostatic pressure at the same time that it is being stressed, much larger strains than are usually encountered can be obtained without fracture, and a large portion of the book is devoted to tests made under pressure. Such practical subjects as wire drawing and extrusion under pressure, and punching under pressure are discussed in detail. The final section of the book gives a comparison of the experimental results with the mathematical theory, and attempts to show how various degrees of theoretical approximation reproduce the facts.

Although for the most part the book is a presentation of experimental data gleaned from numerous experiments, the author uses a certain amount of mathematical analysis to obtain a detailed knowledge of the stress distribution over the specimen. The mathematics is relatively simple, involving only the plasticity equations and the equations of stress equilibrium; however, the solution of these equations is not always straightforward, and in general, the problem reduces to finding a self-consistent set of stresses and strains. It is desirable in such a procedure to have as many experimental data as possible to compare with the theoretical analysis; and in one instance Professor Bridgman uses a very interesting experimental set-up to verify certain assumptions about the strain distribution which he found on theoretical grounds.

Most of the experimental results are on steel, although some data on other metals, glass, carboloy, and minerals are given. The book contains extensive tabular data together with numerous graphs. About twenty photographs of large deformations or of fracture specimens are presented.

The book is certainly a valuable contribution to the field of plastic flow. It will be appreciated by engineers and mathematicians working in the field, and because of the wealth of data, by solid state physicists who are interested in explaining plastic flow phenomena from the atomic point of view. It could easily serve as a text for a person wanting to learn about this interesting

field, except for the fact that the author assumes a certain familiarity with the mathematical theory of plasticity; it is unfortunate, in this reviewer's opinion, that Professor Bridgman did not include an introductory chapter on the subject. The last section of the book, *General Survey*, includes a brief but rather good presentation of the mathematical background, and it is recommended that one not familiar with the subject start his reading here.

JOHN R. REITZ
Los Alamos Scientific Laboratory

The Behavior of Engineering Metals

BY H. W. GILLETT. Pp. 395+xvi, Figs. 51. John Wiley & Sons, Inc., New York, 1951. Price $6.50.

This posthumous book of the famous "Dean of American Metallurgists" compiles a tremendous wealth of factual information and represents concentrated metallurgical know-how. At the same time it claims to be an introduction to facts about metals for the nonspecialist. The reviewer feels, however, that the book will be primarily valued by metallurgists who are familiar with the subject and who enjoy the thinking of a strong personality and of a keen ever-ready mind. In addition, the author presents metallurgical developments up to date and emphasizes many of the current problems, as is also apparent from the references, which cover particularly the papers published in the last few years. It is for this reason, as well as because of the extreme condensation of the material, that the book appears to be less suitable for the beginner and nonmetallurgist. One would expect that a general introduction into engineering materials would primarily offer a concise description of basic and long-established facts and laws rather than the most advanced information, which is very complex and which may be superseded within a comparatively short time. Thus, it is highly recommended that every professional metallurgist add this book to his library of biographical documents. On the other hand, it cannot be expected that it will be equally valuable as a guide for the solution of daily engineering problems.

GEORGE SACHS
Syracuse University

Equivalent Circuits of Electric Machinery

BY GABRIEL KRON. Pp. 278+x. John Wiley & Sons, Inc., New York, 1951. Price $10.00.

In his important researches into the theory of electric machines, Kron has used some of the most advanced concepts of tensor analysis. Among the achievements of his theories has been the discovery of equivalent circuits for electric machines of all sorts. The purpose of this book is not to explain the formidable tensor theory that led Kron to these equivalent circuits, but simply to present the actual circuits. Accordingly tensors are referred to only in general terms, in a short prologue and epilogue, and do not appear explicitly in the main part of the book.

After a preliminary chapter outlining in simple terms the physical picture on which this particular aspect of his theories is based, Kron proceeds to give equivalent circuits for a staggering variety of electric machines. These machines may be at rest, in steady motion, accelerating, or hunting (though hunting is not treated in detail), and they may be interconnected in any manner and under all sorts of load conditions. Basic equivalent circuits are first given for a very general type of machine called the primitive machine, and the equivalent circuits for particular machines are obtained from these by making modifications that correspond to those that would have to be made physically if one wished to convert the idealized primitive machine into the given practical machines. These modifications correspond to tensor transformations in the underlying mathematical theory. In this way a unified approach is obtained to the treatment of all industrial machines.

The equivalent circuits are stationary networks, except for the presence, in some cases, of phase shifting devices. Their stationary character is achieved by an ingenious use of the tensor concept of absolute time derivatives, whereby the basic equivalent circuits for machines in motion are reduced to the same form as those for the machines at rest; only the constants of the circuits are different. The circuits are, as the author stresses, physical models of the machines and thus are capable, when set up, of giving direct numerical information about all principal aspects of the functioning of the machines. Space harmonics are treated by regarding each harmonic as belonging to a separate machine, these separate machines then being interconnected to form the given machine. Time harmonics in unbalanced systems are treated in terms of special rotating reference axes, the equivalent circuits remaining stationary.

As one would expect in a work on such a subject, the book is profusely illustrated. The available space is about evenly divided between text and diagrams, there being over one-hundred and fifty figures, many of the individual figures containing sets of related equivalent circuits.

Though the book is not primarily intended for those whose main interest is in the mathematical details of the underlying tensor equations, numerous references to the original mathematical papers are given in the footnotes. In the text, however, the discussion is largely intuitive and physical, the mathematics being kept extremely simple—just enough to set the stage for the various equivalent circuits and to explain how information may be extracted from them. While this will be particularly appealing to those who are accustomed to thinking in terms of equivalent circuits, the book can be recommended to all who are interested in knowing about or in applying commercially some of the specific practical results that have emerged from Kron's classic researches. As an account of these practical results this notable book is without parallel, and indeed without precedent.

BANESH HOFFMANN
Queens College

General Network Analysis

BY WILBUR R. LEPAGE AND SAMUEL SEELY. Pp. 516+vii, McGraw-Hill Book Company, Inc., New York, 1952. Price $8.00.

The title of this book is somewhat misleading because only electrical circuits whose parameters are linear and constant are considered, and therefore, most of the material given is concerned with a presentation of methods for solving sets of linear differential equations with constant coefficients. Several textbooks on this general subject have been published during the last five years. The purpose of this volume, as the authors state in their preface, is to provide the background for advanced work in electric circuit theory in either power or communications. A previous course in network analysis on the part of the student is presupposed. Although some filter design theory is included, the text deals primarily with analysis and covers the harmonic steady-state behavior of lumped and distributed parameter networks and transient analysis of simple lumped parameter systems. Brief discussions of specialized mathematical subjects of importance in circuit analysis are given. These include matrix algebra, Fourier series and integrals, and the elements of the Laplace transform theory.

The existing texts on electric circuit analysis tend to fall into two distinct categories. The authors of some of these books begin their treatment by obtaining the general solutions of the differential equations that govern the current or potential distribution of the general series or parallel circuits. In these books, the steady-state harmonic oscillations are discussed as special cases of the general solution after the transient effects have disappeared. Other authors determine the particular integrals of the differential equations involved by the method of undetermined coefficients

and thus obtain the periodic parts of the solutions without mentioning the general solution until the latter part of their texts. The authors of this book follow the first plan. The impedance and admittance concepts are introduced in the second chapter by obtaining the particular solution of the linear differential equations involved in a most unorthodox manner which is lacking in mathematical rigor. This procedure is surprising, in view of the fact that the method of the Laplace transform, which is discussed at a later stage, could have been used to obtain the desired solution in a most elegant and concise manner.

Little or no emphasis is placed on the energy concept as it applies to electrical circuits or to a discussion of fundamental circuit parameters, or to that of the most useful circuit analogies of mechanical and acoustical or thermal systems. The physicist, who is not aquainted with the penchant for elaborate terminology possessed by some electrical engineers, may be surprised at the authors' need to introduce the terms, "phasor," "sinor," "copedance," "comittance," and "immitance" (spelled immitance in the index and immittance in chapter six) into such an old and well-established branch of electricity. The attempt to give physical significance to *negative* frequency in chapter six is also somewhat confusing.

It is gratifying to note the emphasis given to the principle of duality and to the use of matrix algebra in the chapters on four terminal networks and polyphase systems. In the opinion of the reviewer, these subjects are extremely useful and important and are not usually included in works on electric circuit analysis.

LOUIS A. PIPES
University of California

The Rolling of Metals

BY L. R. UNDERWOOD. Volume 1. Pp. 344+ix, Figs. 142. John Wiley & Sons, Inc., New York, 1950. Price $6.50.

The technique of rolling metals has been developed industrially as an art to a high state of efficiency. In more than a century and a half, the knowledge of that technique was built up bit by bit by the ways of trial and error. It is only within our own time that a proper scientific approach to the problems of metallic deformation in rolling mills has been fostered. The author in this volume records the extent of recent scientific development and the resulting progress in the technique of rolling metals. He indicates the principal factors involved in the phenomena of rolling and illustrates the manner in which they are related. The author's conclusions were drawn either directly from the quoted experimental work of others or from comparisons of theoretical findings with the results of experiment.

This book covers the fundamental aspects of the subject such as forward slip, friction between rolls and material, and friction hill theories of rolling and methods of calculation of the rolling load based upon them. Due consideration is given to the simplest case of rolling in which ideal conditions are used for the sake of illustration. In this manner the reader is introduced to the forces acting in the region of contact along with their consequences. Then the case is considered in which the resistance of deformation varies along the arc of contact owing to work-hardening in cold rolling and to variation of compression rate in hot rolling, without tension and with front and back tension, either singly or together. Considered next is the added complication of the elastic deformation of the roll in the arc of contact. Finally, the most complicated case is treated in which work-hardening, roll flattening, front tension, and back tension all occur together.

The entire subject is presented in a masterful and scholarly manner. The book should be of considerable value to the practical roller, of especial value to the designer of rolling mills, and also to the group of men who are interested in experimental rolling and the theories of rolling.

C. NUSBAUM
Case Institute of Technology

Other Books Received

The Rise of the New Physics. A. D'ABRO. Vol. I. Pp. 426+ix, Figs. 37 and 24 portraits. Vol. II. Pp. 429–982, Figs. 38–61 and 12 portraits. Dover Publications, New York, 1952. Price $8.00 for the set of two volumes.

Intermediate College Mechanics. DAN EDWIN CHRISTIE. Pp. 454+xiii, Figs. numerous. McGraw-Hill Book Company, Inc., New York, 1952. Price $7.00.

Application of the Electronic Valve in Radio Receivers and Amplifiers. B. G. DAMMERS, J. HAANTJES, J. OTTE, AND H. VAN SUCHTELEN. Pp. 431+xviii, Figs. 343. Philips' Technical Library, Eindhoven, Holland, 1952.

Electrolytic Manganese and Its Alloys. REGINALD S. DEAN. Pp. 257+xx, Figs. 100. The Ronald Press Company, New York, 1952. Price $12.00.

Plastics Molding. JOHN DELMONTE. Pp. 493+vii. John Wiley & Sons, Inc., New York, 1952. Price $9.00.

The Thermodynamics of the Steady State. K. G. DENBIGH. Pp. 103+vii. John Wiley & Sons, Inc., New York, 1952. Price $1.75.

Du Pont—The Autobiography of an American Enterprise. Pp. 138. Published in commemoration of the 150th Anniversary of the founding of E. I. du Pont De Nemours & Company. Distributed by Charles Scribner's Sons, New York, 1952.

The Conduction of Electricity Through Gases. K. G. EMELEUS. Pp. 99+x, Figs. 37. John Wiley & Sons, Inc., New York, 1952. Price $1.50.

Introduction to Electronic Circuits. R. FEINBERG. Pp. 163+xiv. Longmans, Green and Company, New York, 1952. Price $3.50.

The Phase Rule. A. FINDLAY, A. N. CAMBELL, AND N. O. SMITH. Ninth edition. Pp. 512, Figs. 236. Dover Publications, New York, 1952. Price $1.90.

Two New Sciences. GALILEO GALILEI. Translated by Henry Crew and Alfonso de Salvio. Pp. 300+xxi, Figs. 126. Dover Publications, New York, 1952. Price $1.50.

The Aurorae. L. HARANG. Vol. I. Pp. 166, Figs. 152. John Wiley & Sons, Inc., New York, 1951. Price $4.50.

Elements of Thermodynamics and Statistical Mechanics. E. O. HERCUS. Pp. 153+ix, Figs. 28. Cambridge University Press, New York, 1952. Price $3.75.

Transmitting Valves. J. P. HEYBOER AND P. ZIJLSTRA. Pp. 284+xii, Figs. 256. Philips' Technical Library, Eindhoven, Holland.

Short-Wave Radiation Phenomena. AUGUST HUND. Vols. I and II. Pp. 1382+vii, Figs. 394. McGraw-Hill Book Company, Inc., New York, 1952. Price $20.00 for the set of two volumes.

Advanced Theory of Waveguides. L. LEWIN. Pp. 192, Figs. 54. Iliffe & Sons, Ltd., London, or The British Book Centre, Inc., New York, 1951. Price $6.75.

Concepts and Methods of Theoretical Physics. ROBERT BRUCE LINDSAY. Pp. 513+x. D. Van Nostrand Company, Inc., New York, 1951. Price $7.00.

Principles of Biochemistry. BRIAN MASON. Pp. 276+ix, Figs. 42. John Wiley & Sons, Inc., New York, 1952. Price $5.00.

Matter and Motion. JAMES CLERK MAXWELL. Pp. 163, Figs. 17. Dover Publication, Inc., New York, 1952. Price $1.25.

Musical Engineering. HARRY F. OLSON. Pp. 369+vii. McGraw-Hill Book Company, Inc., New York, 1952. Price $6.50.

Photosynthesis and Related Processes. EUGENE I. RABINOWITCH. Vol. II, Part 1. Pp. 1208. Interscience Publishers, Inc., New York, 1951. Price $15.00.

Transient Electric Currents. HUGH K. SKILLING. Second edition. Pp. 361+vii, Figs. 105+7 Oscillograms. McGraw-Hill Book Company, Inc., New York, 1952. Price $6.00.

Directory for the British Glass Industry. Fifth edition. Pp. 415, 8vo. The Society of Glass Technology, Sheffield, England, 1952.

The Organ. WILLIAM LESLIE SUMNER. Pp. 436, Figs. 36+18 Plates. Philosophical Library.

The Science of Flames and Furnaces. M. W. THRING. Pp. 416+xiv. John Wiley & Sons, Inc., New York, 1952. Price $6.50.

Glass and W. E. S. Turner. A volume containing articles by twenty contributors and edited by E. J. GOODING AND E. MEIGH. Pp. 144, Tables 30. The Society of Glass Technology, Sheffield, England, 1952.

Coloured Glasses. W. A. WEYL. Pp. 541, 8vo, Figs. 116. The Society of Glass Technology, Sheffield, England, 1952.

The Planets: Their Origin and Development. HAROLD C. UREY. Pp. 245+xvii. Yale University Press, New Haven, 1952. Price $5.00.

Electromagnetics. ROBERT M. WHITMER. Pp. 270+ix. Prentice-Hall, Inc., New York, 1952. Price $5.00.

Mechanik III: Dynamik der Systems. HANS ZIEGLER. Pp. 306, Figs. 191. Verlag Birkhäuser, Basel, Switzerland, 1952.

Fouriersynthese von Kristallen. WERNER NOWACKI. Pp. 237, Figs. 120. Verlag Birkhäuser, Basel, Switzerland, 1952.

Announcements

Meetings

Industrial Management Society Clinic

The Sixteenth Annual Time and Motion Study and Management Clinic sponsored by the Industrial Management Society will be held on November 5–6–7, 1952, at the Sheraton Hotel, Chicago. Topflight industrial leaders from all over the United States will discuss the latest developments in the fields of time study, motion economy, job evaluation, methods, plant layout, materials handling, and human relations.

Inquiries may be addressed to the Industrial Management Society, 35 East Wacker Drive, Chicago 1, Illinois.

Electron Microscope Society Meets in Cleveland

The Electron Microscope Society will hold its 1952 meeting at the Hotel Statler in Cleveland on November 6, 7, and 8.

Recent Appointments

Dr. C. M. Blair has been named Superintendent of the new chemicals and plastic resins plant to be constructed at Seadrift (Calhoun County) Texas, by the Carbide and Carbon Chemicals Company, a Division of Union Carbide and Carbon Corporation.

John R. Madigan, physicist for the National Bureau of Standards, Washington, D.C., has been appointed an associate physicist by Armour Research Foundation of Illinois Institute of Technology. His appointment as a member of the visual and microwave spectroscopy section was announced by the chairman of the physics department.

Colonel Herbert Milwit has assumed command of the Engineer Research and Development Laboratories, Fort Belvoir, Virginia.

Victor J. Pollock has joined the staff of Consolidated Engineering Corporation, Pasadena, California, as assistant to the Treasurer of the Company. Mr. Pollock was born in Hamilton, Ontario, Canada, and was raised in Springfield, Massachusetts. He has had wide and varied business experiences.

Granvil C. Kyker, Professor and Head of the Department of Biochemistry and Nutrition, School of Medicine, University of Puerto Rico, has been appointed Principal Scientist of the Medical Division of the Oak Ridge Institute of Nuclear Studies. Dr. Kyker succeeds Dr. H. D. Bruner, who left Oak Ridge on July 1 to become Professor and Chairman of the Department of Physiology, Emory University Medical School.

University of Utah Fellowship

Establishment of a postdoctoral fellowship at the University of Utah, Salt Lake City, has been announced by Radioactive Products, Inc., Detroit, Michigan. The fellowship will be for the research on synthesis of radioactive carbon labeled steroids.

Radioactive Products, Inc. manufactures radioactivity detecting and measuring instruments, as well as several specialized labeled C-14 compounds.

Charles B. Jolliffe Scholarship Established at West Virginia University

An undergraduate scholarship for students majoring in pure science or engineering was established today by the Radio Corporation of America at West Virginia University, Morgantown, West Virginia. Brig. General David Sarnoff, Chairman of the Board of RCA, announced that "the award provides an annual grant of $800 and will be known as the Charles B. Jolliffe Scholarship in honor of the Vice President and Technical Director of RCA."

Candidates for the Charles B. Jolliffe Scholarship will be selected by academic officers of West Virginia University and recommended to the RCA Education Committee for consideration. The first recipient of the award will be announced during the Fall of 1952.

Other RCA scholarships are offered, on a competitive basis, to outstanding undergraduate science students at seventeen universities throughout the United States.

Nuclear Announces New Catalog

A new 11-page, 2-color condensed catalog listing Nuclear's latest instruments and radiochemicals for radioisotope applications in the medical, educational, industrial, and research fields is now available on request from Nuclear Instrument and Chemical Corporation, 229 West Erie Street, Chicago 10, Illinois.

Revised Leeds & Northrup Catalog Describes New Speedomax Recorders

To describe how Speedomax Recorders and Controllers are being applied to an increasing variety of uses in industrial production and research, Leeds & Northrup Company has just published a revised, 48-page edition of its Speedomax catalog—*Speedomax Type G Instruments for Measurement and Control.* A copy of Catalog ND46 (1) will be sent upon request. Address Leeds & Northrup Company, 4934 Stenton Avenue, Philadelphia, Pennsylvania.

New Sintered Magnet Catalog Offered by Indiana Steel

Just off the press is the new Catalog No. 12 by The Indiana Steel Products Company. This catalog lists specifications for a range of magnets of special interest to makers of thermostats, meters, switches, and other relatively small devices. It also describes in detail the magnetic and physical characteristics of these powdered metal magnets. Catalog No. 12 is available upon request. Write The Indiana Steel Products Company, Valparaiso, Indiana.

Cannon Electric Revised Bulletin

Cannon Electric has issued a completely revised 32-page Audio-Type Connector Bulletin (No. PO-4-1952). This bulletin is available upon request to the Cannon Electric Advertising Department, P.O. Box 75, Lincoln Heights Station, Los Angeles 31, California.

New Honeywell Literature

Minneapolis-Honeywell Regulator Company has re-issued Bulletin 15–14, "Instruments Accelerate Research." The bulletin now comprises 96 pages and has complete illustrations with pictures of the latest instrument and equipment in the laboratory, analytical, and measurement fields. Inquiries should be addressed to Station 40, Minneapolis-Honeywell Regulator Company, Philadelphia 44, Pennsylvania.